AD-A052 579 AIR WEATHER SERVICE SCOTT AFB ILL F/G 4/2
GUIDE FOR APPLIED CLIMATOLOGY.(U)
NOV 77

UNCLASSIFIED AWS-TR-77-267 NL

1 OF 2
AD
A052579

IFIED

1 OF 2

AD
A052579

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

1

AD-A052579

AWS-TR-77-267

GUIDE FOR APPLIED CLIMATOLOGY

Approved For Public Release;
Distribution Unlimited

PUBLISHED BY
HQ AIR WEATHER SERVICE (MAC)
UNITED STATES AIR FORCE
NOVEMBER 1977

REVIEW AND APPROVAL STATEMENT

This report is approved for public release. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This technical report has been reviewed and is approved for publication.

Ralph M. Koplowitz

RALPH M. KOPLOWITZ
Information Retrieval Officer
Reviewing Officer

FOR THE COMMANDER

Walter S. Burgmann

WALTER S. BURGMANN
Scientific and Technical Information Officer (STINFO)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AWS TR-77-267	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER

4. TITLE *(and Subtitle)*

Guide for Applied Climatology

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Air Weather Service (MAC)
Scott AFB, IL 62225

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Air Weather Service (MAC)
Scott AFB, IL 62225

12. REPORT DATE

November 1977

13. NUMBER OF PAGES

154

14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*

15. SECURITY CLASS. *(of this report)*

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

Formerly published as Air Weather Service Pamphlet 105-2, 1 November 1968.

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Climate, Climatology, Probability, Statistics, Distributions, Regression, Correlation, Extreme Value Analysis, Winds, Time Series

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This publication is a useful guide to some applied climatological practices utilized by AWS meteorologists in support of military plans and operations. It explains the application of statistics and probability techniques to climatological problem-solving, and provides several examples of methods useful in solving recurring requests.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

Preface

"Applied Climatology" can be defined as the scientific analysis of climatological data in the light of its useful application for a practical purpose. This technical report offers the AWS meteorologist a useful guide to applied climatological practices. It is distinctly oriented towards, but not limited to, military applications. It explains the application of statistics and probability to problem-solving, and provides examples of methods useful in solving recurring requests.

Since the US Air Force Environmental Technical Applications Center (USAFETAC) is the organization designated as the facility of Air Weather Service for providing climatic services to military agencies and other authorized organizations, this technical report covers, for the most part, techniques and methods employed by USAFETAC personnel. The examples selected for inclusion allow the reader to follow the practical applications of probability and statistics in solving some of the more frequent requests received by the military climatologist.

This technical report is a reprint and redesignation of AWSP 105-2, 1 November 1968, including Change 1, dated 22 December 1969. Chapters 1 and 2 of that document have been deleted as they are now outdated. No other changes, either editorial or in content, have been made.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 3

PROBABILITY

---

1. Introduction.

a. The applied climatologist is concerned with the application of probability theory in the interpretation of climatological data. The concept of probability involves the notion of prediction of the number of times an attribute, A, will occur in a month, in a year, or in any interval of time.

b. If an event can occur in N mutually exclusive and equally likely ways and if f of these outcomes have an attribute A, then the probability (P) of A is the fraction f/N

(1) $$P(A) = \frac{f}{N}$$

c. The concept of probability, i.e., estimated or empirical probability, as a relative frequency is the most meaningful in applied climatology. As the sample size increases, the observed relative frequency (probability) approaches the true relative frequency in the universe.

Example 1: What is the probability that Washington, D. C. will have a ceiling < 1000 feet and/or visibility < 3 miles in February? The answer depends on the period of record used to determine the probability. Thus, if we used only one year (1950), the probability would be:

$$P(A) = \frac{f}{N} = \frac{121}{672} = 0.180$$

For five years (1950-1954) $$P(A) = \frac{f}{N} = \frac{308}{3384} = 0.091$$

For ten years (1950-1959) $$P(A) = \frac{f}{N} = \frac{889}{6789} = 0.131$$

Because of the larger sample size, the probability determined from ten years of data is most likely nearer the true probability than that given by the data for one or five years.

d. Basic probability theorems are shown below. In addition to the usual theorems, the relationship between the odds of an attribute occurring and the probability of that attribute is given by Equation (3).

. The probability of a "certain" attribute is 1.

. The probability of an "impossible" attribute is 0.

. The probability of the attribute A is $0 \leq P(A) \leq 1$.

. If A and a are complementary attributes, then

$$P(A) = 1 - P(a) \tag{2}$$

. The odds for the attribute A are m to n, if and only if

$$P(A) = \frac{m}{m + n} \tag{3}$$

Using Equation (3), the following Odds versus Probability of attribute A are obtained:

Odds for A	P(A)
1 to 1	1/2
2 to 1	2/3
3 to 5	3/8
1 to 2	1/3

2. Statistics of Attributes (Set Theory).

a. Statistical analysis deals with quantitative data that arise in two different ways. In the first case, the observer notes the occurrence or non-occurrence of some attribute in a series of observations. The methods applicable to this type are referred to as "statistics of attributes" [55]. In the second case, the observer notes or measures the actual magnitude of some variable. The methods applicable to these cases are referred to as "statistics of variables."

b. Attributes are divided into two distinct classes. Letters A, B, and C will be used to denote the several attributes. All members that possess attribute A, B, or C will be termed Class A, Class B, or Class C. All members not possessing attribute A, B, or C will be termed Class a, b, or c (lower case letters mean not A, not B, and not C).

c. The number of observations assigned to a class is called the class frequency. Class frequencies of attributes are designated by order number depending on the number of attribute classes included. For example, AB, aB, bC are classes of the second order; ABc, aBc, AbC are classes of the third order. For three attributes there are 3^3, or 27, distinct frequencies. For n attributes there are 3^n distinct frequencies. A class frequency can be expressed in terms of class frequencies of higher order, as

classes).

(ABC)	225	(aBC)	675
(ABc)	225	(aBc)	900
(AbC)	600	(abC)	1350
(Abc)	375	(abc)	3150

(A) = (ABC) + (ABc) + (AbC) + (Abc) = 1425

(B) = (ABC) + (ABc) + (aBC) + (aBc) = 2025

(AB) = (ABC) + (ABc) = 450

N = (ABC) + (ABc) + (AbC) + (Abc) + (aBC) + (aBc) + (abC) + (abc) = 7500

The complete results are:

N = 7500	(B) = 2025	(AB) = 450	(BC) = 900
(A) = 1425	(C) = 2850	(AC) = 825	(ABC) = 225

e. A <u>fundamental set</u> is a set of 2^n class frequencies that are well defined and distinct. The fundamental set specifies the whole data. The positive class frequencies are a fundamental set, i.e., N, (A), (B), (C), (AB), (AC), (BC), (ABC). The following example shows how, if given the positive class frequencies of Example 2, all of the class frequencies can be determined.

<u>Example 3</u>: Given a fundamental set, e.g., the eight positive class frequencies of Example 2, find all the class frequencies.

N = 7500;	(A) = 1425;	(B) = 2025;	(C) = 2850;
(AB) = 450;	(AC) = 825;	(BC) = 900;	(ABC) = 225

We have:

(AB) = (ABC) + (ABc)

450 = 225 + (ABc)

(ABc) = 225

Similarly from (AC) and (BC)

(AbC) = 600

(aBC) = 675

also

$$(abC) = (bC) - (AbC)$$
$$= (C) - (BC) - (AbC)$$
$$= 2850 - 900 - 600$$
$$= 1350$$

Similarly, we have

$$(Abc) = 375$$
$$(aBc) = 900$$

Finally

$$(abc) = (bc) - (Abc)$$
$$= (c) - (Bc) - (Abc)$$
$$= N - (C) - \{(B) - (BC)\} - (Abc)$$
$$= 7500 - 2850 - 1125 - 375$$
$$= 3150$$

f. Class frequencies that have been or might have been observed in the same population may be said to be <u>consistent</u> with one another. For this consistency to exist, the necessary and sufficient condition is that no ultimate class frequency be negative [55].

<u>This condition must exist</u>	<u>or this frequency will be negative</u>
Two Attributes	
$(AB) \geq 0$	(AB)
$(AB) \geq (A) + (B) - N$	(ab)
$(AB) \leq (A)$	(Ab)
$(AB) \leq (B)$	(aB)
Three Attributes	
$(ABC) \geq 0$	(ABC)
$\geq (AB) + (AC) - (A)$	(Abc)
$\geq (AB) + (BC) - (B)$	(aBc)
$\geq (AC) + (BC) - (C)$	(abC)
$\leq (AB)$	(ABc)
$\leq (AC)$	(AbC)
$\leq (BC)$	(aBC)
$\leq (AB) + (AC) + (BC) - (A) - (B) - (C) + N$	(abc)

g. It is possible to draw inferences from data that are otherwise insufficient for calculating all class frequencies by use of the limits set forth in the preceding paragraph.

<u>Example 4</u>: Given: Airfield A, open 80% of the time;

Airfield B, open 95% of the time:

Find limits to the percentage of time A and B are simultaneously open, i.e., (AB).

$$(AB) \geqq 0 \qquad (A) = 80$$

$$(AB) \geqq 80 + 95 - 100 = 75 \qquad (B) = 95$$

So, both airfields are open not less than 75% or more than 80% of the time.

h. Attributes may be <u>independent</u> of each other, <u>completely associated</u>, <u>partially associated</u>, or <u>completely disassociated</u>.

(1) Independence. If the occurrence of one attribute gives no information as to the occurrence of the other attribute, they are said to be independent of each other. Thus, with independent attributes A and B, we can expect the same proportion of A's among the B's as among the not B's, i.e., $(AB)/(B) = (Ab)/(b)$. This relationship can be seen in the 2 × 2 table below:

TABLE 1

Independent Attributes.

Attribute	B	b	Total
A	(AB)	(Ab)	(A)
a	(aB)	(ab)	(a)
Total	(B)	(b)	N

The proportion of A's among B's is the same as that in the universe, i.e.,

$$\frac{(AB)}{(B)} = \frac{(AB) + (Ab)}{(B) + (b)} = \frac{(A)}{N}$$

$$\frac{(AB)}{(A)} = \frac{(B)}{N} \; ; \quad (AB) = \frac{(A)(B)}{N}$$

(4) $$\frac{(AB)}{N} = \frac{(A)}{N} \times \frac{(B)}{N} \quad \text{or} \quad (AB) = \frac{(A)(B)}{N}$$

Equation (4) is the fundamental rule for independence.

(2) Attributes A and B are associated if they appear together in a greater number of cases than is expected of independent attributes, e.g., $(AB) > (A)(B)/N$; if $(AB) < (A)(B)/N$, A and B are disassociated. For complete association (AB) must be equal to (A) or (B), whichever is smaller. For complete disassociation (AB) must be equal to zero or $(A) + (B) - N$, whichever is greater.

(3) The association between A and B in subuniverses, defined by C and c, are called partial associations. A and B are positively associated in the population C if $(ABC) > (AC)(BC)/C$, and negatively associated in the converse case.

(4) The coefficient of complete association (a) must not be dependent upon the size of the sample or frequency of the attributes; and (b) must be convenient to use, i.e., the coefficient is zero when independent, +1 when completely associated, and -1 when completely disassociated. A simple coefficient that satisfies these two requirements is

$$(5) \qquad Q = \frac{(AB)(ab) - (Ab)(aB)}{(AB)(ab) + (Ab)(aB)}$$

An example of the use of Equation (5) is shown below:

Example 5: The frequency of halos and subsequent precipitation was observed. The following 2 × 2 table summarizes the observations:

TABLE 2

Summary of Halos and Subsequent Precipitation.

	Precipitation Within 48 Hours		Total
	Yes B	No b	
Halo A	497	149	646
No Halo a	819	819	1638
Total	1316	968	2284

Using Equation (5)

$$Q = \frac{(497)(819) - (149)(819)}{(497)(819) + (149)(819)} = 0.54$$

(5) For partial associations, the coefficient of association Q is defined as:

$$(6) \qquad Q_{AB.C} = \frac{(ABC)(abC) - (AbC)(aBC)}{(ABC)(abC) + (AbC)(aBC)}$$

(6) An example of the use of the coefficient of association equations is shown below:

Example 6: Let us suppose that the following percentage frequency of attributes has been observed.

	C	c	Total
AB	12	12	24
Ab	28	8	36
aB	3	18	21
ab	7	12	19
Total	50	50	100

The total association between A and B from Equation (5) is

$$Q_{AB} = \frac{(24)(19) - (36)(21)}{(24)(19) + (36)(21)}$$

$$= -0.25$$

This indicates a negative association exists between A and B in the total universe. Now we consider the partial association between A and B in subuniverse of C and c by Equation (6)

$$Q_{AB.C} = \frac{(12)(7) - (28)(3)}{(12)(7) + (28)(3)} = 0$$

$$Q_{AB.c} = \frac{(12)(12) - (8)(18)}{(12)(12) + (8)(18)} = 0$$

This indicates that A and B are independent of each other in both subuniverses.

i. Manifold classification [55] is the division of a population according to an attribute A into a number of classes. The fundamental principles of manifold classification are similar to, but more complicated than, those for a 2×2 classification. Manifold classification according to two attributes A and B gives a contingency table where the classification of the A's is s-fold and that of the B's t-fold. There will be $s \times t$ classes of the type A_mB_n, as represented by the following table:

TABLE 3

Manifold Contingency.

Attribute	B_1	B_2		B_t	Totals
A_1	(A_1B_1)	(A_1B_2)	...	(A_1B_t)	(A_1)
A_2	(A_2B_1)	(A_2B_2)	...	(A_2B_t)	(A_2)
⋮					
A_s	(A_sB_1)	(A_sB_2)	...	(A_sB_t)	(A_s)
Totals	(B_1)	(B_2)	...	(B_t)	N

j. Association in a contingency table can be examined by reducing it in a number of ways to a 2 × 2 table, but the procedure is very long and tedious if s and t are large. In practice, we usually want a coefficient that will summarize the general nature of the dependence. Two such coefficients are Pearson's "coefficient of mean-square contingency" and Tschuprow's "coefficient of contingency." These coefficients are discussed below:

(1) If A's and B's are completely independent in the universe, then for all values of i's and j's

(7) $$(A_iB_j) = \frac{(A_i)(B_j)}{N} = (A_iB_j)_o$$

where $(A_iB_j)_o$ is the expected frequency. If A and B are not independent, then the difference d_{ij} equals:

(8) $$d_{ij} = (A_iB_j) - (A_iB_j)_o$$

We define the quantity χ^2 as follows:

(9) $$\chi^2 = \sum_i \sum_j \left[\frac{d_{ij}^2}{(A_iB_j)_o}\right]$$

<u>NOTE</u>: χ^2 must be calculated from the actual number of observations (not percentage frequencies) in each class and the total number of observations N must be known, and the expected frequency $(A_iB_j)_o$ should be at least five for any class.

From the above, we have the "coefficient of mean-square contingency" as given by Karl Pearson:

$$(10) \qquad C = \left[\frac{\chi^2}{(N + \chi^2)}\right]^{1/2}$$

(2) The coefficient shown in Equation (10) has one serious disadvantage, i.e., it never reaches 1 as a limit. To remedy this, Tschuprow proposed the coefficient T defined by:

$$(11) \qquad T^2 = \frac{\chi^2}{N[(s-1)(t-1)]^{1/2}}$$

The coefficient of Equation (11) varies between 0 and 1 when s = t.

<u>Example 7</u>: Are the following field conditions independent of the given synoptic hours?

TABLE 4

Summary of Field Conditions.

Hours	Field Conditions			Totals
	Closed	Inst	Contact	
06	55	59	218	332
12	4	33	298	335
18	7	19	307	333
Total	66	111	823	1000

Degrees of freedom (df)

= (no. of rows minus 1) (no. of columns minus 1)

= (3-1)(3-1) = 4

Using Equation (9)

$$\chi^2 = \sum_i \sum_j \left[\frac{d_{ij}^2}{(A_iB_j)_o}\right] = 114.4$$

Using Equation (10)

$$C = \left[\frac{\chi^2}{(N + \chi^2)}\right]^{1/2} = 0.32$$

Using Equation (11)

$$T = \left[\frac{\chi^2}{N[(s-1)(t-1)]^{1/2}}\right]^{1/2} = 0.24$$

where s = t = 3

Hence, field conditions are significantly dependent on synoptic hours (06, 12, 18) and not likely by chance.

k. In previous subsections, we were concerned with the intersection or joint occurrence of attributes A and B, i.e., the occurrence of both A and B. One other concept is important: the union of attributes A and B, which is the occurrence of either A or B, or both. The intersection of A and B is denoted as AB; the union of A and B is denoted as A∪B. These two concepts can be seen in Figure 2, where the intersection of A and B is represented by the shaded area and the union of A and B is shown by the area enclosed by the heavy line. The union of A and B = Ab + AB + aB = A + B - AB = 1 - ab.

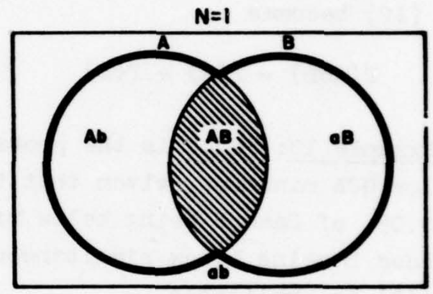


Figure 2. The Intersection and Union of Attributes A and B.

l. The probability that at least one of two attributes, i.e., A and/or B, occurs is equal to the sum of the probabilities of each event minus the probability that both events occur simultaneously,

(12) $P(A \cup B) = P(A) + P(B) - P(AB)$

<u>Example 8</u>: An attack on a coastal installation is being planned. Troops and equipment can be delivered to the area by air, sea, or both. What is the probability of success as a result of weather effects? Defining A as an event favorable for delivery of troops by sea and B by air, climatological information provides the following fundamental set:

$P(A) = .50 \quad P(B) = .60 \quad P(AB) = .25$

The probability of at least one mode favorable is

$P(A \cup B) = .50 + .60 - .25 = .85$

m. The probability that at least one of two attributes occurs is also

equal to one minus the probability that neither attribute occurs, i.e.,

(13) $P(A \cup B) = 1 - P(ab)$

Example 9: What is the probability of < 1000 feet ceiling and/or < 2 miles visibility at Harmon in January. From the "D" summary for Harmon, the probability of ≥ 1000 feet ceiling and ≥ 2 miles visibility equals 0.83. This gives

$$P(A \cup B) = 1 - 0.83 = 0.17$$

n. If attributes A and B are mutually exclusive, i.e., $P(AB) = 0$, Equation (12) becomes

(14) $P(A \cup B) = P(A) + P(B)$

Example 10: What is the probability that Base A and/or Base B are below GCA minimums, given that the probability of Base A being below is 0.05, of Base B being below minimum is 0.03, and of both Base A and Base B being below simultaneously is zero? Since $P(AB) = 0$, Equation (14) is applicable.

$$P(A \cup B) = P(A) + P(B) = 0.05 + 0.03 = 0.08$$

o. The definition of conditional probability can be stated as follows: The probability of attribute A, given that attribute B has occurred, is equal to the probability of both A and B divided by the probability of B.

(15) $P(A/B) = \frac{P(AB)}{P(B)}$

Equation (15) is undefined if $P(B) = 0$.

Example 11: Going back to Example 8, what is the probability of success by sea, given that air delivery is unfavorable.

$$P(A/b) = \frac{P(Ab)}{P(b)} \qquad P(A) = P(AB) + P(Ab) = .50$$

$$P(Ab) = P(A) - P(AB)$$

$$P(Ab) = .50 - .25 = .25$$

$$P(A/b) = \frac{.25}{.40} = .625 \qquad P(b) = 1 - P(B)$$

$$P(b) = 1 - .60 = .40$$

So, if it is known that air delivery is unfavorable, the probability

of success by sea is increased from .50 to .625.

Example 12: What is the probability that Base A will be above alternate minimums (ceiling $\geq$ 1000 feet and visibility $\geq$ 3 miles), given that Base B is below GCA minimums (ceiling < 200 feet and/or visibility < 1/2 mile)? Again we use Equation (15):

$$P(A/B) = \frac{P(AB)}{P(B)}$$

From a special summary that gives simultaneous conditions for Bases A and B, P(AB) = .05, and from the "D" summary for Base B, P(B) = .06. Then

$$P(A/B) = \frac{.05}{.06} = .83$$

This is an important and often useful application of the conditional probability equation.

p. It should be noted that all general probability theorems are also valid for conditional probabilities with respect to any given attribute. For the *probability of occurrence* of either A or B, or both, given attribute C has occurred, we have

$$P(A \cup B/C) = P(A/C) + P(B/C) - P(AB/C) \tag{16}$$

or in a form more easily used

$$P(A \cup B/C) = \frac{P(AC) + P(BC) - P(ABC)}{P(C)} \tag{17}$$

Example 13: What is the probability that either Base A or Base B or both will be above alternate minimums (ceiling $\geq$ 1000 and visibility $\geq$ 3 miles), given that Base C is below GCA minimums (ceiling < 200 feet and/or visibility < 1/2 mile)? From a special summary that gives simultaneous conditions for Bases A, B, and C, we find

$$P(AC) = .04 \qquad P(BC) = .07 \qquad P(ABC) = .02$$

And from the "D" summary for Base C

$$P(C) = .10$$

then

$$P(A \cup B/C) = \frac{.04 + .07 - .02}{.10} = \frac{.09}{.10} = .90$$

q. The probability of the attribute A and the conditional probability of A, given the attribute B has occurred, are generally unequal, but they can be equal, i.e.,

(18) $P(A/B) = P(A)$

This means that knowing attribute B has occurred does not change the probability of the attribute A. Therefore, the attribute A is independent of the attribute B. The attribute B must have positive probability. And if

$P(A) > 0$ and $P(B) > 0$

we can rewrite Equation (15)

(19) $P(AB) = P(A)\ P(B/A) = P(B)\ P(A/B)$

and like Equation (18)

(20) $P(B/A) = P(B)$

When Equations (18) and (20) hold, Equation (19) becomes

(21) $P(AB) = P(A)\ P(B) = P(B)\ P(A)$

r. Attributes A and B are said to be independent if, and only if, Equation (21) holds, i.e., the probability that both A and B occur is the product of the probability that A occurs and the probability that B occurs. If the probability of the attribute A can be assumed to be independent of attribute B, then Equation (21) can be used to give joint occurrence of A and B.

<u>Example 14</u>: What is the probability of ceiling ≥ 200 feet and visibility ≥ 1/2 mile at Elmendorf and simultaneously, a ceiling of ≥ 1000 feet and visibility ≥ 2 miles at Eielson in January? If we assume the two attributes A (Elmendorf) and B (Eielson) to be independent, the "D" summaries give $P(A) = .96$ and $P(B) = .80$, so

$P(AB) = (.96)(.80) = .77$

A special simultaneous summary for Elmendorf and Eielson gave $P(AB) = .76$, so the assumption of independence was a good one. Similarly, computations using Equation (21) can be extended to several attributes:

(22) $P(ABC) = P(A)\ P(B)\ P(C)$

s. Yule and Kendall [55] note that, when Equation (21) holds for attributes A and B, it will also hold for $P(Ab)$, $P(aB)$, and $P(ab)$. However, when

Equation (22) holds for A, B, and C, as in Example 15 below, no conclusion can be drawn about P(ABc), P(AbC), ... P(abc) without further information.

Example 15: Given:

Base A < 1000 feet and/or 3 miles, $P(A) = .19$

Base B < 1000 feet and/or 3 miles, $P(B) = .27$

Base C < 1000 feet and/or 3 miles, $P(C) = .38$

What is the probability all three are simultaneously below 1000/3? If we assume that attributes A, B, and C are independent, then

$$P(ABC) = (.19)(.27)(.38) = .02$$

A special simultaneous summary gave $P(ABC) = .03$, so the assumption of independence was quite good.

3. Permutations and Combinations.

a. The number of permutations of r attributes selected from n given distinct attributes, where attention is paid to the order in which the attributes are selected, is given by:

$$(23) \qquad (n)_r = P(n,r) = \frac{n!}{(n - r)!}$$

Example 16: Let (A, B, C, D) be a set of four air bases. How many wind factors will be necessary to cover all the routes between each of the four bases? The wind factor from A to B is different than that from B to A, and so on for the other bases. Consequently, as we select two of the air bases at a time, attention must be paid to the order in which they are selected. This gives:

$$P(4,2) = \frac{4!}{(4 - 2)!} = \frac{4!}{2!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1} = 4 \times 3 = 12$$

So there are 12 wind factors to calculate. The 12 pairs or 12 permutations are:

AB	AC	AD	BC	BD	CD
BA	CA	DA	CB	DB	DC

b. The required number of combinations of r attributes selected from n given distinct attributes, where the order in which the attributes are selected is not important, is given by:

$$(24) \qquad \frac{P(n,r)}{r!} = \binom{n}{r} = \frac{n!}{r!(n - r)!}$$

Example 17: For the four air bases in Example 16 let P(A), P(B), P(C), P(D) be the probabilities that airfields A, B, etc., are closed. What are all the possible combinations of the air bases being closed? If we select zero, one, two, three, and then four airfields from the set of four and pay no attention to the order of selection, we form all the possible combinations as shown below:

(a) Zero, $\binom{4}{0} = \frac{4!}{0!4!} = 1$ where $0! = 1$

None of the airfields are closed.

P(abcd)

(b) One, $\binom{4}{1} = \frac{4!}{1!3!} = \frac{4.3.2.1}{1.3.2.1} = 4$

P(A), P(B), P(C), P(D)

(c) Two, $\binom{4}{2} = \frac{4!}{2!2!} = \frac{4.3.2.1}{2.1.2.1} = 6$

P(AB), P(AC), P(AD), P(BC), P(BD), P(CD)

(d) Three, $\binom{4}{3} = \frac{4!}{3!1!} = \frac{4.3.2.1}{3.2.1.1} = 4$

P(ABC), P(ABD), P(ACD), P(BCD)

(e) Four, $\binom{4}{4} = \frac{4!}{4!0!} = 1$

P(ABCD)

So we have 1 + 4 + 6 + 4 + 1 = 16 well-defined and distinct probabilities that will specify the whole data of four attributes A, B, C, and D.

4. Binomial Distribution.

a. The binomial distribution can be used to find probabilities and percentiles when only the mean value is known. However, it can be used only when two alternatives or outcomes are possible, i.e., the attribute occurs or does not occur on an individual trial. Also, there are two important assumptions associated with the binomial distribution method: (1) the trials (N) are independent, and (2) the probability (P) remains the same throughout the trials.

b. The binomial probability function for a given p and n, with k as the number of successes, is given by

(25) $$b(k; n,p) = \binom{n}{k} p^k q^{n-k} \qquad k = 0, 1, \ldots n$$

where p is the probability of occurrence and q = 1 - p is the probability of nonoccurrence. If we denote the number of successes k in n trials by S_n, then

$$(26) \qquad b(k; n,p) = P(S_n = k) = \binom{n}{k} p^k q^{n-k}$$

Equation (26) represents the k^{th} term of the binomial expansion

$$(27) \qquad (q + p)^n = b(0; n,p) + b(1; n,p) + b(2; n,p) + \ldots + b(n; n,p) = 1$$

Example 18: Washington, D. C. can expect 6.4 days with thunderstorms in July, i.e., p = 6.4/31 = 0.206. If we assume that the occurrence of a thunderstorm on a given day is independent of whether a thunderstorm occurred the day before, what is the probability of at least four days with thunderstorms during the month of July? The probability is:

$$P(S_{31} \geq 4) = 1 - P(S_{31} \leq 3)$$

$$P(S_{31} \geq 4) = 1 - [P(S_{31} = 0) + P(S_{31} = 1) + P(S_{31} = 2) + P(S_{31} = 3)]$$

$$P(S_{31} = 0) = \binom{31}{0} (.206)^0 \ (.794)^{31} = \frac{31!}{0!(31-0)!} (1) \ (.794)^{31}$$

$$= \text{anti-}\ell n \ [31 \ \ell n \ (.794)] = 0.0008$$

$$P(S_{31} = 1) = \binom{31}{1} (.206)^1 \ (.794)^{30} = \frac{31!}{1!(31-1)!} (.206) \ (.794)^{30}$$

$$= 31 \ (.206) \ [\text{anti-}\ell n \ (30 \ \ell n \ .794)] = 0.0070$$

$$P(S_{31} = 2) = \binom{31}{2} (.206)^2 \ (.794)^{29}$$

$$= \frac{31!}{2!(31-2)!} (.206)^2 \ [\text{anti-}\ell n \ (29 \ \ell n \ .794)] = 0.0255$$

$$P(S_{31} = 3) = \binom{31}{3} (.206)^3 \ (.794)^{28}$$

$$= \frac{31!}{3!(31-3)!} (.206)^3 \ [\text{anti-}\ell n \ (28 \ \ell n \ .794)] = 0.0618$$

The probability of at least four days with thunderstorms is one minus the sum of these four probabilities.

$$P(S_{31} \geq 4) = 1.0000 - (0.0008 + 0.0070 + 0.0255 + 0.0618)$$

$$= 1.0000 - 0.0950 = 0.905$$

The probability of four or more days with thunderstorms, based on the 16 observed cases in the past 19 years, is 0.85. This difference,

0.905 - 0.850 = 0.055 or 5.5%, probably is due in part to the persistence factor that will be discussed in the next paragraph.

c. The binomial probability function may be difficult to calculate unless binomial tables are available, but if N is larger than 25 and p is between 0.25 and 0.75, the binomial distribution may be approximated by the normal distribution. If the normal approximation is used, probabilities are easily determined from probability paper (Figure 3) or normal tables using the mean and standard deviation of the binomial distribution. The mean of the binomial distribution is:

(28) $$M_b = Np$$

where N is the number of independent trials, and p is the probability of occurrence of the attribute. The standard deviation of the binomial distribution is:

(29) $$\sigma_b = \sqrt{Npq}$$

Where, as before, p is the probability of occurrence and q = 1 - p is the probability of nonoccurrence of the attributes. Meteorological phenomena tend to have a persistence effect, e.g., if a phenomenon (thunderstorm, rain, fog, etc.) has occurred on a given day, it is more likely to occur the next day than if it had not occurred on the given day. Consequently, the observed standard deviation usually will be somewhat higher than that calculated from Equation (29). This difference, the observed minus the binomial standard deviation, was not large enough to affect significantly the probabilities of the attributes considered in Example 18. However, Brooks and Carruthers [11] gives persistence factors for various phenomena, which when multiplied by σ_b^2 may give a better estimate of the standard deviation, i.e.,

$$\sigma = \sqrt{S_n \times \sigma_b^2}$$

Selected factors from the referenced text are:

Phenomenon	Persistence Factor(s)
Thunder	1.4
Rain	1.7
Fog	2.0
Snow	2.3

Example 19: Colorado Springs, Colorado can expect 14.6 days with thunderstorms during the month of August. What is the probability of at least 18 days with thunderstorms? If we assume independence, the

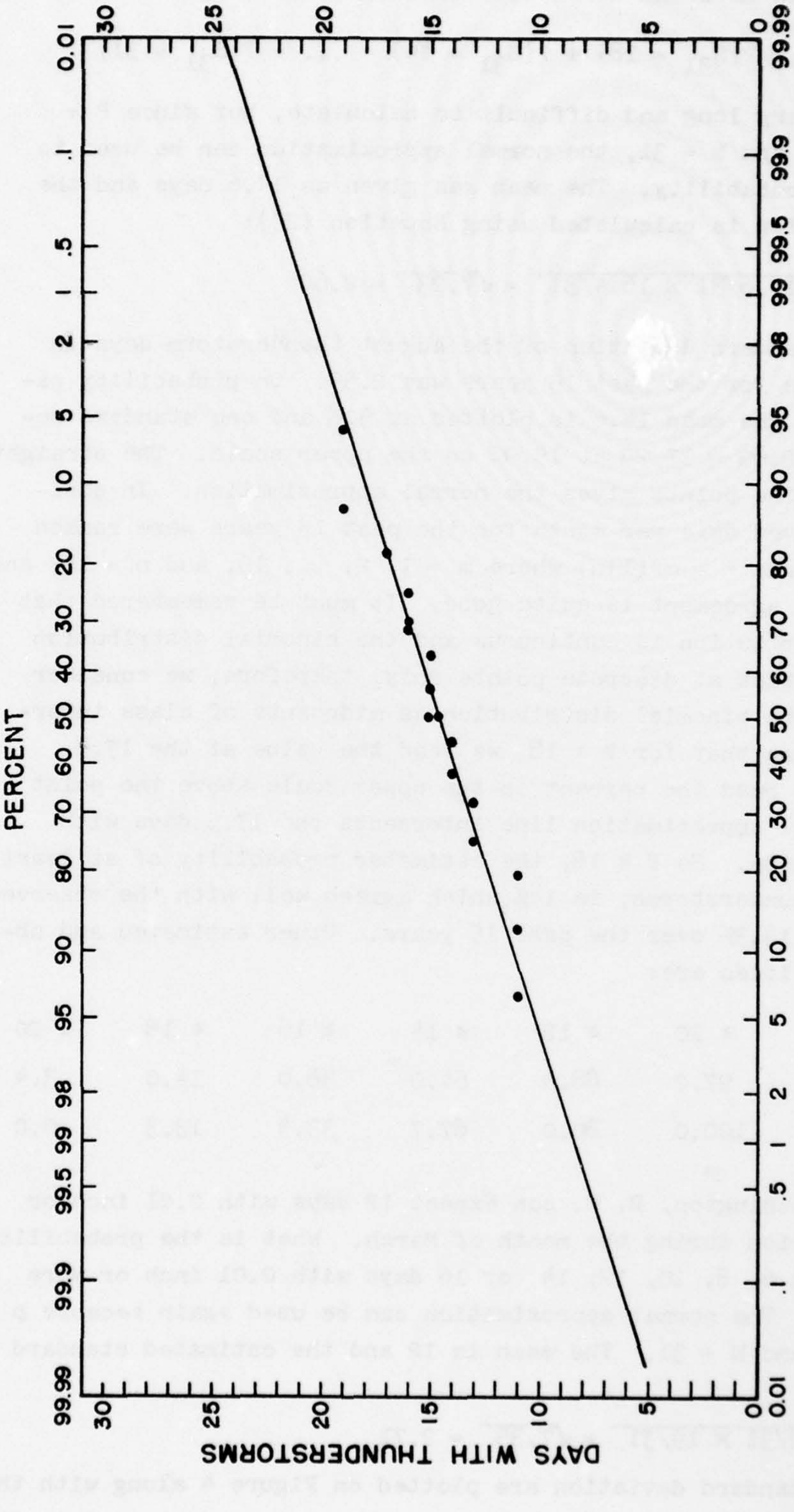


Figure 3. The Normal Approximation with M = 14.6 and P = 0.470.

probabilities can be found using Equations (26) and (27).

$$P(S_{31} \geq 18) = P(S_{31} = 18) + P(S_{31} = 19) + \ldots + P(S_{31} = 31)$$

This would be very long and difficult to calculate, but since P = 14.6/31 = 0.470 and N = 31, the normal approximation can be used to determine the probability. The mean was given as 14.6 days and the standard deviation is calculated using Equation (29):

$$\sigma_b = \sqrt{31 \times 14.6/31 \times 16.4/31} = \sqrt{7.23} = 2.69$$

The observed standard deviation of the August thunderstorm days in Colorado Springs for the past 15 years was 2.59. On probability paper (Figure 3), the mean 14.6 is plotted at 50% and one standard deviation 14.6 + 2.69 = 17.29 at 15.9% on the upper scale. The straight line through these points gives the normal approximation. In addition, the observed days per month for the past 15 years were ranked (using the formula P = m/(1+n) where m = 1, 2, ... 15, and n = 15) and plotted and the agreement is quite good. It must be remembered that the normal distribution is continuous and the binomial distribution gives probabilities at discrete points only; therefore, we consider the points of the binomial distribution as midpoints of class intervals. This means that for $P \geq 18$ we read the value at the 17.5 point, i.e., we read the percent on the upper scale above the point where the normal approximation line intersects the 17.5 days with thunderstorms line. So $P \geq 18$, the estimated probability of at least 18 days with thunderstorms, is 14%, which agrees well with the observed probability of 13.3% over the past 15 years. Other estimated and observed probabilities are:

Days per Month	≧ 10	≧ 12	≧ 14	≧ 16	≧ 18	≧ 20
Estimated	97.2	88.0	64.0	38.0	14.0	3.4
Observed	100.0	80.0	67.7	33.3	13.3	0.0

Example 20: Washington, D. C. can expect 12 days with 0.01 inch or more precipitation during the month of March. What is the probability of no more than 6, 8, 10, 12, 14, or 16 days with 0.01 inch or more precipitation? The normal approximation can be used again because p = 12/31 = 0.387 and N = 31. The mean is 12 and the estimated standard deviation is:

$$\sigma_b \sqrt{31 \times 12/31 \times 19/31} = \sqrt{7.35} = 2.71$$

The mean and standard deviation are plotted on Figure 4 along with the

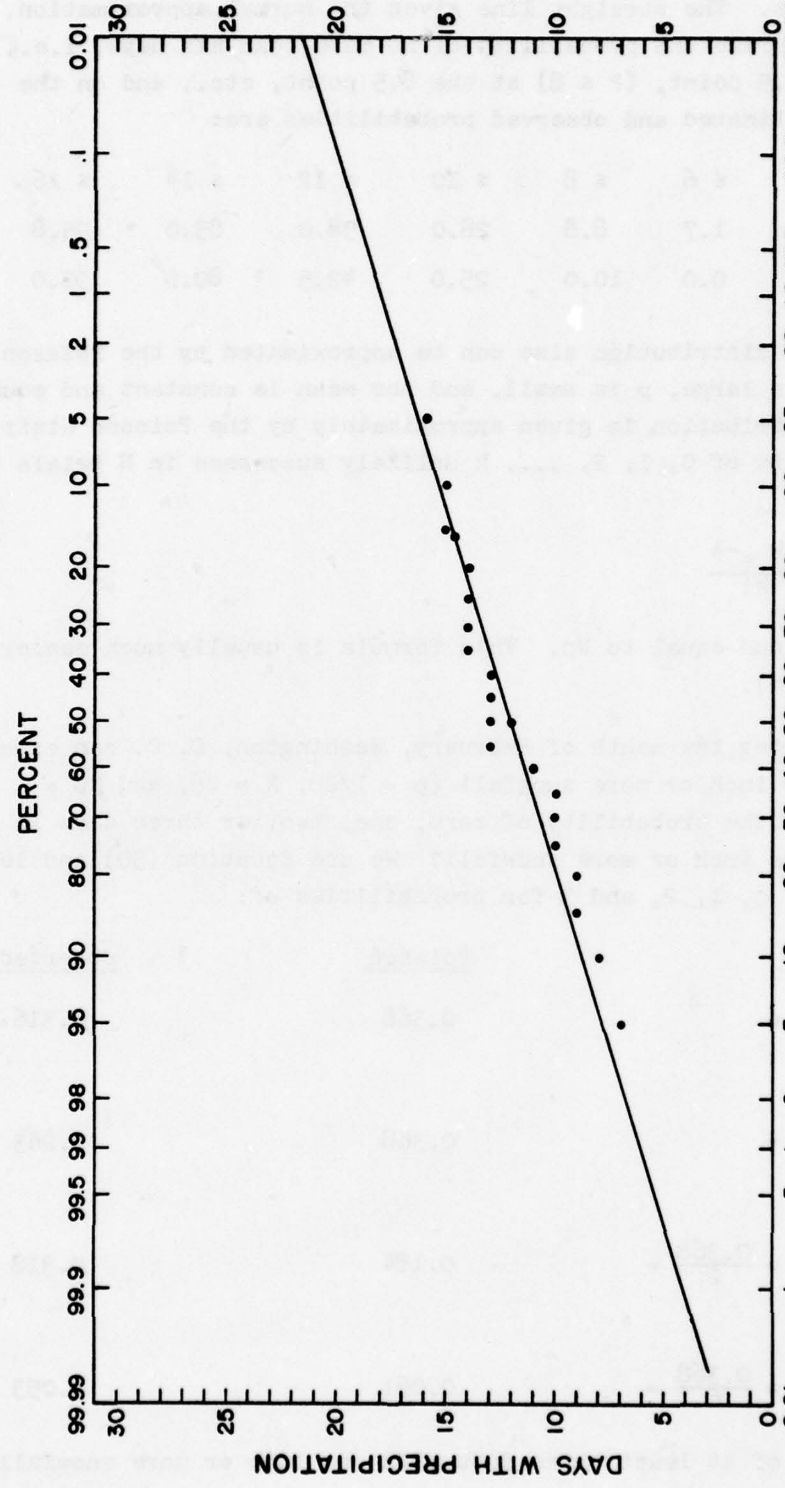


Figure 4. The Normal Approximation with M = 12.0 and P = 0.387.

19 observed cases. The straight line gives the normal approximation. In this case, we read the probability of no more than six days, i.e., $(P \leq 6)$ at the 6.5 point, $(P \leq 8)$ at the 8.5 point, etc., and on the lower scale. Estimated and observed probabilities are:

Days per Month	≤ 6	≤ 8	≤ 10	≤ 12	≤ 14	≤ 16
Estimated	1.7	8.8	28.0	58.0	83.0	95.8
Observed	0.0	10.0	25.0	42.5	80.0	95.0

d. The binomial distribution also can be approximated by the Poisson distribution. When N is large, p is small, and the mean is constant and equal to Np, the binomial distribution is given approximately by the Poisson distribution. The probability of 0, 1, 2, ..., k unlikely successes in N trials is approximated by:

$$(30) \qquad P(\lambda, k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where λ is the mean and equal to Np. This formula is usually much easier to use than Equation (26).

<u>Example 21</u>: During the month of February, Washington, D. C. can expect one day with one inch or more snowfall ($p = 1/28$, $N = 28$, and $Np = \lambda = 1$). What is the probability of zero, one, two, or three days in February with one inch or more snowfall? We use Equation (30) and let k take values of 0, 1, 2, and 3 for probabilities of:

	Poisson	Observed
Zero Days:		
$P(1,0) = \frac{1^0 e^{-1}}{0!} =$	0.368	0.316
One Day:		
$P(1,1) = \frac{1^1 e^{-1}}{1!} =$	0.368	0.263
Two Days:		
$P(1,2) = \frac{1^2 e^{-1}}{2!} = \frac{0.368}{2} =$	0.184	0.318
Three Days:		
$P(1,3) = \frac{1^3 e^{-1}}{3!} = \frac{0.368}{6} =$	0.061	0.053

The probability of at least three days with one inch or more snowfall is:

$$P(S_{28} \geq 3) = 1 - [P(S_{28} = 0) + P(S_{28} = 1) + P(S_{28} = 2)]$$

$$= 1 - (0.368 + 0.368 + 0.184) = 1 - 0.920 = 0.08$$

In the past 19 years, three or more days with one inch or more snowfall has occurred twice, i.e., $p = 2/19 = 0.105$, which agrees well with the Poisson approximated value of 0.08.

Chapter 4

STATISTICS

1. Introduction.

a. Statistics, as applied to meteorology, involves the analysis of past weather data and allows the analyst to draw conclusions about the behavior of similar weather data in the future. For short-range forecasting, the meteorologist relies heavily on a review of conditions immediately preceding the forecast period. On the other hand, the meteorological statistician summarizes past data according to time and space without reference to immediate conditions. Such summaries have their greatest utility in long-range forecasting.

b. In meteorology, statistical methods are used to obtain, analyze, and present numerical weather data. Methods range from the most elementary descriptive devices to extremely complicated mathematical procedures. The meteorological statistician analyzes a series of observations in such a way as to get the best information from them, and applies the results of his analyses to specific questions about climate.

2. Frequency Distributions.

a. The meteorological observation network generates an enormous quantity of data each year. Before the analyst can begin to study these data, he must organize them in a manner that readily points out their major features and simplifies future computations. A method that is frequently used by the analyst is to sort observed data into separate classes. The numerical width of each class is called the class interval and is the difference between the upper mathematical limit of the class and the lower mathematical limit of the class. The midpoint of the class interval is defined as the class mark. The class frequency is the number of observations falling in the class interval. The class percentage frequency is defined as follows:

$$\frac{\text{Class Frequency}}{\text{Total Number of Observations}} \times 100$$

It is desirable, but not mandatory, that all classes have the same class interval. A rule of thumb for determining the number of classes into which these data should be divided is to use ($5 \log_{10} N$) classes where N is the total number of observations. Table 5 is a summary of mean monthly temperatures for Washington National Airport; it will be used in examples throughout

this chapter. A tabular display of data sorted into classes is called a <u>frequency distribution</u>; Table 6 is such a display of the data shown in Table 5.

TABLE 5

Mean Monthly Temperature in January at Washington National Airport.

Year	Temp (°F)	Rank	Temp (°F)	Year	Temp (°F)	Rank	Temp (°F)
1943	36	1	29	1953	41	11.5	36
1944	37	2	30	1954	36	11.5	36
1945	31	3.5	31	1955	35	13.5	37
1946	37	3.5	31	1956	35	13.5	37
1947	42	5	33	1957	33	15	38
1948	29	6	34	1958	34	16	39
1949	43	8.5	35	1959	35	17.5	41
1950	48	8.5	35	1960	38	17.5	41
1951	39	8.5	35	1961	30	19	42
1952	41	8.5	35	1962	35	20	43
				1963	31	21	48

TABLE 6

Frequency Distribution.

Class	Class Mark	Lower Math. Limit	Upper Math. Limit	Class Freq.	Percent Freq.
29-31	30	28.6	31.5	4	19.0
32-34	33	31.6	34.5	2	9.6
35-37	36	34.6	37.5	8	38.0
38-40	39	37.6	40.5	2	9.6
41-43	42	40.6	43.5	4	19.0
44-46	45	43.6	46.5	0	0.0
47-49	48	46.6	49.5	1	4.8
$\sum$	--	--	--	21	100.0

No. of Classes = $5 \log_{10} 21 = 5 \times 1.322 = 6.61 \simeq 7$

Observed Range = $48 - 29 = 19$°F

Class Interval = $19 \div 7 = 2.7 \simeq 3$°F

b. Often the analyst is interested in the frequency (or percentage frequency) of observations with numerical value less than or equal to a specific value. The frequency distribution may be readily converted into a <u>cumulative</u>

frequency distribution by summing the frequencies from the lowest value to the highest value. Table 7 is a cumulative frequency distribution of the frequencies shown in Table 6. A sample reading from Table 7 shows that 28.6% of the time the mean temperature is less than or equal to 34.5°F.

TABLE 7

Cumulative Frequency Distribution.

Class	Upper Math. Limit	Class Freq.	Cumul. Freq.	Percent Freq.	Cumul. Percent Freq.
29-31	31.5	4	4	19.0	19.0
32-34	34.5	2	6	9.6	28.6
35-37	37.5	8	14	38.0	66.6
38-40	40.5	2	16	9.6	76.2
41-43	43.5	4	20	19.0	95.2
44-46	46.5	0	20	0.0	95.2
47-49	49.5	1	21	4.8	100.0

c. A graphical presentation of a frequency distribution is called a histogram. An important feature of a histogram is that each observation is represented by a unit of area. The area representing each class is directly proportional to the class frequency. Figure 5 is a histogram of the frequency distribution of Table 6.

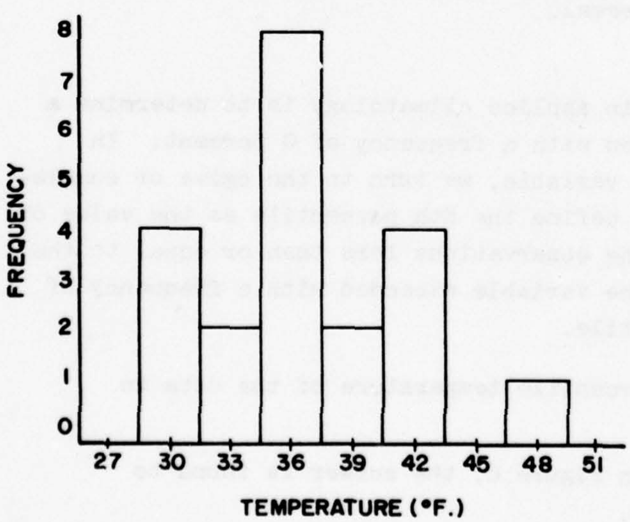


Figure 5. Histogram. When class intervals are equal, the height of the bar is equal to the class frequency.

d. A graphical presentation of the cumulative frequency distribution is called an <u>ogive</u>. The ordinate of an ogive is equal to the frequency (or percent frequency) of observations having values less than or equal to a specified value. This means the ordinate is proportional to the area of a histogram located to the left of the abscissa value. Figure 6 is an ogive of the cumulative frequencies of Table 7.

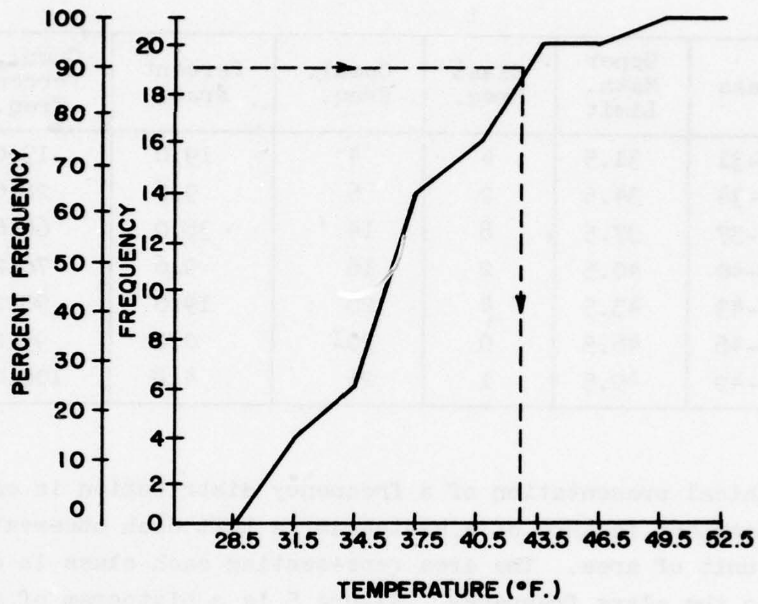


Figure 6. Ogive. The ordinate is plotted at the upper mathematical limit of each class interval.

e. An approach frequently used in applied climatology is to determine a value of the variable that is exceeded with a frequency of Q percent. In order to determine this value of the variable, we turn to the ogive or cumulative frequency distribution. Let us define the Kth percentile as the value of the variable that has K percent of the observations less than or equal to the specified value. The magnitude of the variable exceeded with a frequency of Q percent is then the (1-Q)th percentile.

<u>Example 22</u>: What is the 90th percentile temperature of the data in Table 5?

<u>Method A</u>: Using the ogive in Figure 6, the answer is found to be 42.6°F (see dashed lines).

Method B: Referring to the cumulative frequency distribution of Table 7, use linear interpolation wit^r 1 the approximate class interval (in this example the 41-43°F interval).

$$\frac{43.5 - 40.5}{43.5 - P_{90}} = \frac{95.2 - 76.2}{95.2 - 90} = \frac{19.0}{5.2} = 3.66$$

$$43.5 - P_{90} = \frac{43.5 - 40.5}{3.66} = \frac{3.0}{3.66} = 0.82$$

$$P_{90} = 43.5 - 0.82 = 42.7°F$$

The 90th percentile (P_{90}) temperature is 42.7°F. This result can be restated to indicate that 10% of the observations will have a temperature in excess of 42.7°F.

3. Measures of Central Value.

a. Central value refers to the location of the center of the distribution of observations. There are three of these measures: the median, the mode, and the mean.

b. The median is the middle value of the observations when they are arranged in ranked order (see Table 5). A look at Table 5 shows that the median is 36°F. The median is also the 50th percentile value. The ogive in Figure 6 shows the 50th percentile (median) is 36.15°F. Note the difference between the median as obtained from Table 5 and from the ogive. The difference occurs because in the ranked-order table the temperature is recorded in discrete values, whereas in the ogive we have treated temperature as a continuous variable.

c. The mode of a frequency distribution is that value of the variable for which the frequency is a maximum. A mode then is also the most probable value of the variable. There may be one or more modes to a frequency distribution. A look at the frequency distribution in Table 6, or the histogram of Figure 5, shows the following modes (Table 8):

TABLE 8

Modes.

Temperature	Frequency	Type Mode
30°F	4	Secondary
36°F	8	Primary
42°F	4	Secondary
48°F	1	Tertiary

d. The third and most useful of all measures of central value is the arithmetic mean or the mean. The value of the mean lies in its use in the mathematical theory of statistics. Assume a discrete variate X may assume the values $x_1, x_2, x_3, \ldots, x_N$, where each value has the corresponding probability $p_1, p_2, p_3, \ldots, p_N$. We define the expected value of X as:

$$E(X) = p_1x_1 + p_2x_2 + p_3x_3 + \ldots + p_Nx_N = \sum_{i=1}^{N} p_ix_i$$

The <u>mean</u> is defined as the expected value of X and is shown as E(X) or $\overline{X}$. This is the probabilistic definition of the mean. Simple manipulation can soon transform E(X) into the well-known layman's definition of the mean. Given:

$$E(X) = \sum_{i=1}^{N} p_ix_i$$

From probability theory, we note:

$$p_i = \frac{f_i}{\sum\limits_{i=1}^{N} f_i}$$

then

$$E(X) = \sum_{i=1}^{N} \frac{f_ix_i}{\sum\limits_{i=1}^{N} f_i}$$

but

$$\sum_{i=1}^{N} f_i = N$$

the number of observations, so:

$$E(X) = \frac{1}{N} \sum_{i=1}^{N} f_ix_i$$

If we assume $f_i = 1$ for all individual x_i's, we find

$$E(X) = \frac{1}{N}(x_1 + x_2 + \ldots + x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i$$

The last equation is the layman's definition of the mean or average. Let $Y = aX + b$, then it may be shown that $E(Y) = E(aX + b) = a\,E(X) + b$. This implies that if a constant is added (or subtracted) to all values of X, the mean of the new variable is the constant plus (or minus) the mean of X. Further, if all values of X are multiplied by a constant, the mean of the new variable is equal to the product of the constant and the mean of X. Therefore, there are three procedures available for computing the mean:

$$\text{(31)} \quad \overline{X} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots + f_Nx_N}{f_1 + f_2 + f_3 + \ldots + f_N} = \frac{1}{N}\sum_{i=1}^{N} f_ix_i$$

$$\text{(32)} \quad \overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\text{(33)} \quad \overline{X} = b + \frac{\sum_{i=1}^{N} f_i\,(x_i - b)}{\sum_{i=1}^{N} f_i} = b + \frac{\sum_{i=1}^{N} f_i\,(x_i - b)}{N}$$

NOTE: When the observations are grouped in classes, the x_i's of Equations (31) and (33) are the class marks of each class.

Example 23: Given the data in Table 9, compute the mean using Equations (31), (32), and (33).

Using Equation (31):

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} f_ix_i = \frac{1}{21} \times 766 = 36.5$$

Using Equation (32):

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \doteq \frac{1}{21} \times 766 = 36.5$$

Using Equation (33):

$$\overline{X} = b + \frac{\sum_{i=1}^{N} f_i\,(x_i - b)}{N} = 35 + \frac{31}{21} = 35 + 1.5 = 36.5$$

TABLE 9

Information Extracted from Table 5.

x_i	f_i	$f_i x_i$	x_i-35	f_i (x_i-35)
29	1	29	-6	-6
30	1	30	-5	-5
31	2	62	-4	-8
31	--	--	--	--
33	1	33	-2	-2
34	1	34	-1	-1
35	4	140	0	0
35	--	--	--	--
35	--	--	--	--
35	--	--	--	--
36	2	72	+1	+2
36	--	--	--	--
37	2	74	+2	+4
37	--	--	--	--
38	1	38	+3	+3
39	1	39	+4	+4
41	2	82	+6	+12
41	--	--	--	--
42	1	42	+7	+7
43	1	43	+8	+8
48	1	48	+13	+13
$\sum$ 766	21	766	--	31

Example 24: Given the frequency distribution of Table 10, below, compute the mean using Equation (31).

TABLE 10

Temperature Information from Table 5.

Class	Mark (°F)	Freq.	$f_i x_i$
29-31	30	4	120
32-34	33	2	66
35-37	36	8	288
38-40	39	2	78
41-43	42	4	168
44-46	45	0	0
47-49	48	1	48
$\sum$	--	21	768

Using Equation (31):

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{N} f_i x_i = \frac{1}{21} \times 768 = 36.6°F$$

NOTE: The difference between 36.6°F obtained in this example and 36.5°F found in Example 23 is the result of all observations not falling on class marks of the different classes.

4. Measures of Dispersion.

a. The mean by itself does not provide a clear picture of the distribution. An important feature of a distribution is the extent to which the observed values spread out from the mean, i.e., the dispersion. There are three measures of dispersion:

(1) Range, the difference between the largest and the smallest observed values,

(2) Mean Deviation, the average of the absolute value of the deviations of the observed values from the mean, and

(3) Standard Deviation, the measure of the dispersion of the observed values about their arithmetic mean.

b. Returning to the concept of expectation, variance is defined as $E\,[(X - \overline{X})^2]$ and denoted as "s^2," where "s" is the standard deviation of the sample. Simple manipulation of the defining equation of the variance yields more-workable forms of the equation:

$$(34) \qquad s^2 = \text{Var } X = E\,[(X - \overline{X})^2] = \frac{1}{N}\left[\sum_{i=1}^{N} (X_i - \overline{X})^2\right]$$

also

$$E\,[(X - \overline{X})^2] = E\,(X^2 - 2\overline{X}\,X + \overline{X}^2) = E\,(X^2) - 2\overline{X}\,E(X) + \overline{X}^2$$

$$= E\,(X^2) - 2\overline{X}^2 + \overline{X}^2 = E\,(X^2) - \overline{X}^2$$

but

$$E\,(X^2) = \sum_{i=1}^{N} \frac{f_i x_i{}^2}{N} \qquad \text{and} \qquad \overline{X}^2 = \left[\frac{\sum_{i=1}^{N} (f_i x_i)}{N}\right]^2$$

so

$$s^2 = \sum_{i=1}^{N} \frac{f_i x_i{}^2}{N} - \left[\frac{\sum_{i=1}^{N} (f_i x_i)}{N}\right]^2$$

(35) $$s^2 = \frac{1}{N}\left\{ \sum_{i=1}^{N} f_i x_i{}^2 - \frac{1}{N}\left[\sum_{i=1}^{N} (f_i x_i)\right]^2 \right\}$$

If $f_i = 1$ for all individual i's, then

(36) $$s^2 = \frac{1}{N}\left[\sum_{i=1}^{N} X_i{}^2 - \frac{1}{N}\left(\sum_{i=1}^{N} X_i\right)^2\right]$$

Letting $Y = aX + b$, we have $\overline{Y} = a\overline{X} + b$ and can determine the variance of Y, Var (Y):

$$\text{Var }(Y) = E\,[(Y - \overline{Y})^2] = E\,[(aX + b - a\overline{X} - b)^2] = E\,[a^2\,(X - \overline{X})^2]$$

$$= a^2\,E\,[(X - \overline{X})^2] = a^2\,\text{Var }(X)$$

Thus, adding (or subtracting) a constant to X does not change the variance of X. However, multiplying X by a constant multiplies the variance of X by the constant squared. For example, consider Equation (35):

(37) $$s^2 = \frac{1}{N}\left\{ \sum_{i=1}^{N} f_i x_i{}^2 - \frac{1}{N}\left[\sum_{i=1}^{N} (f_i x_i)\right]^2 \right\}$$

$$= \frac{1}{N}\left\{ \sum_{i=1}^{N} f_i\,(x_i - A)^2 - \frac{1}{N}\left[\sum_{i=1}^{N} f_i\,(x_i - A)\right]^2 \right\}$$

If Equations (34), (35), (36), and (37) are multiplied by N/N-1, the resulting equations give unbiased estimates of the population variance. For computational purposes, any of the following equations can be used to find the variance and standard deviation of our sample:

(38) $$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} (X_i - \overline{X})^2\right]$$

(39) $$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} f_i\,(X_i - \overline{X})^2\right]$$

$$(40) \qquad s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} X_i^{\,2} - \frac{1}{N}\left(\sum_{i=1}^{N} X_i\right)^2\right]$$

$$(41) \qquad s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} f_iX_i^{\,2} - \frac{1}{N}\left(\sum_{i=1}^{N} f_iX_i\right)^2\right]$$

$$(42) \qquad s^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{N} (X_i - A)^2 - \frac{1}{N}\left[\sum_{i=1}^{N} (X_i - A)\right]^2\right\}$$

$$(43) \qquad s^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{N} f_i\,(X_i - A)^2 - \frac{1}{N}\left[\sum_{i=1}^{N} f_i\,(X_i - A)\right]^2\right\}$$

Example 25: Given the data in Table 11 and the results of Example 24, compute the variance and the standard deviation using Equations (38), (41), and (43).

NOTE: Example 23 gives $\overline{X} = 36.5$ and $\sum_{i=1}^{N} f_iX_i = 766$

Using Equation (38):

$$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} (X_i - \overline{X})^2\right] = \frac{1}{20}\,[441.25] = 22.1°F$$

Using Equation (41):

$$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} f_iX_i^{\,2} - \frac{1}{N}\left(\sum_{i=1}^{N} f_iX_i\right)^2\right] = \frac{1}{20}\left[28,382 - \frac{1}{21}\,(766)^2\right]$$

$$= \frac{1}{20}\left[28,382 - \frac{586,756}{21}\right] = \frac{1}{20}\,[441.24] = 22.1°F$$

Using Equation (43):

$$s^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{N} f_i\,(X - A)^2 - \frac{1}{N}\left[\sum_{i=1}^{N} f_i\,(X_i - A)\right]^2\right\}$$

$$= \frac{1}{20}\left\{487 - \frac{1}{21}\,[31]^2\right\} = \frac{1}{20}\left[487 - \frac{961}{21}\right] = \frac{1}{20}\,[441.24] = 22.1°F$$

TABLE 11

Temperature Statistics Derived from Table 5.

X_i	X_i^2	f_i	f_iX^2	$X_i-\overline{X}$	$(X_i-\overline{X})^2$	X_i-35	$(X_i-35)^2$	$f_i(X_i-35)$	$f_i(X_i-35)^2$
29	841	1	841	-7.5	56.25	-6	36	-6	36
30	900	1	900	-6.5	42.25	-5	25	-5	25
31	961	2	1922	-5.5	30.25	-4	16	-8	32
31	961	--	--	-5.5	30.25	--	--	--	--
33	1089	1	1089	-3.5	12.25	-2	4	-2	4
34	1156	1	1156	-2.5	6.25	-1	1	-1	1
35	1225	4	4900	-1.5	2.25	0	0	0	0
35	1225	--	--	-1.5	2.25	--	--	--	--
35	1225	--	--	-1.5	2.25	--	--	--	--
35	1225	--	--	-1.5	2.25	--	--	--	--
36	1296	2	2592	-0.5	.25	+1	1	2	2
36	1296	--	--	-0.5	.25	--	--	--	--
37	1369	2	2738	+0.5	.25	+2	4	4	8
37	1369	--	--	+0.5	.25	--	--	--	--
38	1444	1	1444	+1.5	2.25	+3	9	3	9
39	1521	1	1521	+2.5	6.25	+4	16	4	16
41	1681	2	3362	+4.5	20.25	+6	36	12	72
41	1681	--	--	+4.5	20.25	--	--	--	--
42	1764	1	1764	+5.5	30.25	+7	49	7	49
43	1849	1	1849	+6.5	42.25	+8	64	8	64
48	2304	1	2304	+11.5	132.25	+13	169	13	169
$\sum$ 766	--	21	28382	-0.5	441.25	--	--	+31	487

The standard deviation is

$$s = +\sqrt{22.1} = 4.7$$

Note that $\sum_{i=1}^{N} (X_i - \overline{X})$, the sum of the deviations about the mean, should be zero. This can be proved as follows:

$$\sum_{i=1}^{N} (X_i - \overline{X}) = \sum_{i=1}^{N} X_i - N\,\overline{X} \qquad \text{but} \qquad N\,\overline{X} \equiv \sum_{i=1}^{N} X_i$$

therefore

$$\sum_{i=1}^{N} (X_i - \overline{X}) = \sum_{i=1}^{N} X_i - \sum_{i=1}^{N} X_i \equiv 0$$

However, the sum in column 5 of Table 11 does not equal zero because the mean of $\overline{X}$ is rounded off to 1/10°F.

Example 26: Compute the variance and standard deviation from grouped data using the frequency distribution of Table 12 and Equations (41) and (43).

Using Equation (41):

$$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} f_i X_i^{\,2} - \frac{1}{N}\left(\sum_{i=1}^{N} f_i X_i\right)^2\right] = \frac{1}{20}\left[28{,}548 - \frac{1}{21}(768)^2\right]$$

$$= \frac{1}{20}\left[28{,}548 - \frac{1}{21}(589{,}824)\right] = \frac{1}{20}[461] = 23.05 = 23.1°F$$

Using Equation (43):

$$s^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{N} f_i(X_i - A)^2 - \frac{1}{N}\left[\sum_{i=1}^{N} f_i(X_i - A)\right]^2\right\}$$

$$= \frac{1}{20}\left[585 - \frac{1}{21}(-51)^2\right] = \frac{1}{20}\left[585 - \frac{1}{21}(2601)\right]$$

$$= \frac{1}{20}[461] = 23.05 = 23.1°F$$

Therefore

$$s = +\sqrt{23.1} = 4.8$$

TABLE 12

Temperature Frequency Data Derived from Table 6.

Class Mark	Class Freq.	X_1	X_1^2	f_1X_1	$f_1X_1^2$	X_1-39	$(X_1-39)^2$	$f_1(X_1-39)$	$f_1(X_1-39)^2$
30	4	30	900	120	3600	-9	81	-36	324
33	2	33	1089	66	2178	-6	36	-12	72
36	8	36	1296	288	10368	-3	9	-24	72
39	2	39	1521	78	3042	0	0	0	0
42	4	42	1764	168	7056	+3	9	12	36
45	0	45	2025	0	0	+6	36	0	0
48	1	48	2304	48	2304	+9	81	9	81
$\sum$	21	--	--	768	28548	--	--	-51	585

c. Differences in computed values of the variance and the standard deviation, as given in Examples 25 and 26, are due to the fact that, in Example 26, grouped data were used and all observations do not lie exactly on the class mark of each class interval. The use of grouped data always produces an overestimate of the standard deviation. This may be compensated for by Sheppard's correction.

(44) $$s_c = +\sqrt{s^2 - \frac{i^2}{12}}$$

where s^2 is the variance and i is the class interval.

Example 27: The standard deviation given in Example 26 is corrected as follows:

$$s_c = +\sqrt{23.1 - \frac{(3)^2}{12}} = \sqrt{23.1 - 0.75} = \sqrt{22.35} = 4.7°F$$

d. A standardized, or often referred to as normalized, variable can be defined as a variable with a mean of zero and a standard deviation of one. Such a variable is:

(45) $$Z = \frac{X - \bar{X}}{s}$$

where

$$Z_i = \frac{X_i - \bar{X}}{s}$$

To show that $E(Z) = 0$ and $Var(Z) = 1$

$$\bar{Z} = E(Z) = E\left[\frac{X - \bar{X}}{s}\right] = \frac{1}{s} E[X - \bar{X}] = \frac{1}{s}(0) = 0$$

$$Var(Z) = E\left[\left(\frac{X - \bar{X}}{s} - \bar{Z}\right)^2\right] = E\left[\left(\frac{X - \bar{X}}{s} - 0\right)^2\right] = E\left[\left(\frac{X - \bar{X}}{s}\right)^2\right]$$

$$= \frac{1}{s^2} E[(X - \bar{X})^2] = \frac{s^2}{s^2} = 1$$

5. Universe and Sample.

a. The universe is defined as all measurements of a meteorological parameter that have or could have been taken and can or will be taken at a particular geographical location (or locations). The sample is defined as all measurements of the meteorological parameter that have been taken and are available to the analyst for study. The sample is, therefore, a restricted subset of the universe.

b. There are three statistical distributions connected with the concept of universe and sample: the distribution of the universe, the distribution of the sample, and the distribution of the sample statistics of all possible samples of Sample Size N. Also, most statistical procedures are based on two assumptions: the universe (or population) is infinite, and the individual measurements in the sample are chosen in a random, independent manner. While the universe of all meteorological parameters is infinite, the sample size of many parameters may be quite small. The accepted practice in most climatological studies is to use a time series of observations, i.e., all consecutive values available are used. This practice, in many cases, violates the second assumption noted above, for such observations may be autocorrelated.

c. Autocorrelation coefficients are ordinary correlation coefficients within a time series in which the correlated values are a constant interval apart. The autocorrelation for any lag τ can be estimated from the sample by:

$$(46) \qquad r_\tau = \frac{1}{(N-\tau)\, s_x^{\,2}} \sum_{i=1}^{N-\tau} (X_i - \overline{X})(X_{i+\tau} - \overline{X})$$

An autocorrelation different from zero means that observations are not independent and that some statistical tests in common use are not valid for time-series data. However, an autocorrelation of zero does not necessarily imply that the observations are independent. It is important that the analyst be aware of the problems of using an autocorrelated time series. Special methods have been proposed to take into consideration the fact that a time series may not be an independent random sample.

d. In general, the mean of a sample that consists of independent random observations is a good estimate of the mean of the universe. The variance of this same sample (when computed using formulas given in paragraph 4 of this chapter) is a good estimate of the variance of the universe. Paragraph 7 of this chapter provides a method of determining the confidence limits of the sample mean and variance for either independent or autocorrelated data.

e. The design of an experiment, or the method of choosing a sample, is of importance in all meteorological problems. Paragraph 5a of this chapter contains the definition of the sample. However, in some experiments, although the finite population is known, one must fabricate the random sample. A table of random numbers is useful for this purpose. It consists of numbers selected in a manner similar to drawing numbered slips of paper from a hat. The table is sufficiently large so that all numbers from zero through nine appear with about the same frequency. By combining numbers, the analyst can obtain pairs, three numbers at a time, and so forth. In using such a table, the analyst

should be sure to enter the table in a random manner. One method is to close your eyes and place your finger on a page of the table. The digits under your finger can be used, or these digits may be used to locate others. For example, placing your finger over 2913 can be interpreted as using the digits in the 29th row of column 13.

(1) In a recent problem involving dry periods at an air base, it was necessary to resort to a table of random numbers to fabricate a random sample. A dry period for this problem was defined as a period during which no precipitation greater than 0.10 inch occurred on any one day. Unfortunately, consecutive daily rainfall amounts were not available, so the lengths of dry periods could not be determined. However, since the average number of days per month with precipitation greater than 0.10 inch was available, a table of random numbers was used to select wet and dry days of each month.

(2) Using January as an example, we knew there were 31 items in the universe and from the record, we knew the average January had eight wet days. Consequently, it was necessary to use two-digit columns from the random number table. If the number selected from the table was 31 or less, we called the corresponding day a wet day until eight wet days were selected. If the number was greater than 31 or duplicated a selected number, it was skipped. Following this procedure, these days were selected as wet days in January: 1, 5, 8, 9, 10, 17, 20, and 25. Using the same method, wet days for each month of the average year were selected. Knowing the wet days, it was an easy matter to determine the dry days and runs of consecutive dry days in an average year.

(3) Did use of random numbers give a valid sample? Referring to the actual data, the probability of a wet day throughout the year was found to be 0.24 and the probability of a dry day to be 0.76. Assuming no persistence, the following data for an average year were computed (see Table 13). Notice that four of the comparisons agree exactly, the numbers generally become less as N increases for both actual and computed values, and, in most cases, numerical values are comparable. These similarities indicate the fit of the run of dry days determined by use of the random number table is fairly good and should be tested mathematically. A χ^2 (Chi-square) test indicates that a worse fit would occur by chance as often as one time in two (see paragraph 6g). Consequently, wet days selected by use of the table of random numbers represent a valid sample.

TABLE 13

Check of Results Derived by Random Numbers Techniques.

	Actual		By Random Numbers
N Days	Runs of N or More Dry Days	Runs of Exactly N Dry Days	Runs of Exactly N Dry Days
1	67	16	12
2	51	13	14
3	38	9	9
4	29	7	7
5	22	5	3
6	17	4	4
7	13	3.2	4
8	9.8	2.4	4
9	7.4	1.8	1
10	5.6	1.3	0
11	4.3	1.0	1
12	3.3	0.8	1
13	2.5	0.6	2
14	1.9	0.5	0
15	1.4	0.3	0
16	1.1	0.3	0
17	0.8	0.2	1

6. The Normal Distribution.

a. The normal distribution is probably the most important frequency distribution in meteorological statistics. The normal curve is symmetrical, bell-shaped, and extends infinitely in both positive and negative directions. The normal curve that best fits a particular sample of data is defined as that curve with the same area, mean, and standard deviation as the sample. The equation of the normal curve that best fits the sample is:

$$(47) \qquad Y = \frac{Ni}{s\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x - \bar{x}}{s}\right)^2}$$

where the ordinate Y equals the height of the curve for a given x, $\bar{x}$ is the mean, and s the standard deviation of the sample, N is equal to the total number of observations, i is the class interval, and the area under the normal curve is equal to the area of the histogram of the sample.

b. In order to make calculations easier, Equation (47) can be normalized by letting $z = (x - \bar{x})/s$; then z is normally distributed with mean equal to zero and standard deviation equal to one. The normalized equation is:

(48) $$Y_z = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2}$$

where the area under the curve is equal to one. The curve defined by Equation (48) is called the normal density function and values of Y_z are given in Table 14.

TABLE 14

Ordinates of the Normal Curve.

Y_z	0.	1.	2.	3.
.0	.3989	.2420	.0540	.0044
.1	.3970	.2179	.0440	.0033
.2	.3910	.1942	.0355	.0024
.3	.3814	.1714	.0283	.0017
.4	.3683	.1497	.0224	.0012
.5	.3521	.1295	.0175	.0009
.6	.3332	.1109	.0136	.0006
.7	.3123	.0940	.0104	.0004
.8	.2897	.0790	.0079	.0003
.9	.2661	.0656	.0060	.0002

If the ordinate values as given by Equation (47) for the values of x at the center of each interval (the class mark) are plotted, the smooth curve drawn through these points is the theoretical normal curve of best fit to the data sample.

Example 28: Fit a normal curve to the histogram of data given in Table 5. $N = 21$, $i = 3$, $\bar{x} = 36.5$, and $s = 4.7$.

TABLE 15

Computation of Ordinate Values.

Class	Class Mark	$x - \bar{x}$	$\frac{x - \bar{x}}{s}$	Y_z	$Y = \frac{Ni}{s} Y_z$	Observed Frequency
20-22	21	-15.5	3.2979	.0017	0.0228 or 0.0	0
23-25	24	-12.5	2.6596	.0118	0.1581 or 0.2	0
26-28	27	-9.5	2.0213	.0518	0.6941 or 0.7	0
29-31	30	-6.5	1.3830	.1532	2.0529 or 2.1	4
32-34	33	-3.5	0.7446	.3022	4.0495 or 4.0	2
35-37	36	-0.5	0.1064	.3968	5.3171 or 5.3	8
38-40	39	2.5	0.5319	.3463	4.6407 or 4.6	2
41-43	42	5.5	1.1702	.2012	2.6961 or 2.7	4
44-46	45	8.5	1.8085	.0788	1.0560 or 1.1	0
47-49	48	11.5	2.4469	.0202	0.2707 or 0.3	1
50-52	51	14.5	3.0852	.0035	0.0469 or 0.0	0

c. Figure 7 shows the theoretical normal curve, with the same mean and standard deviation, fitted to the data given in Table 15. The fit does not appear to be good, but a histogram can be misleading because the frequencies depend on the selection of the classes and the class interval.

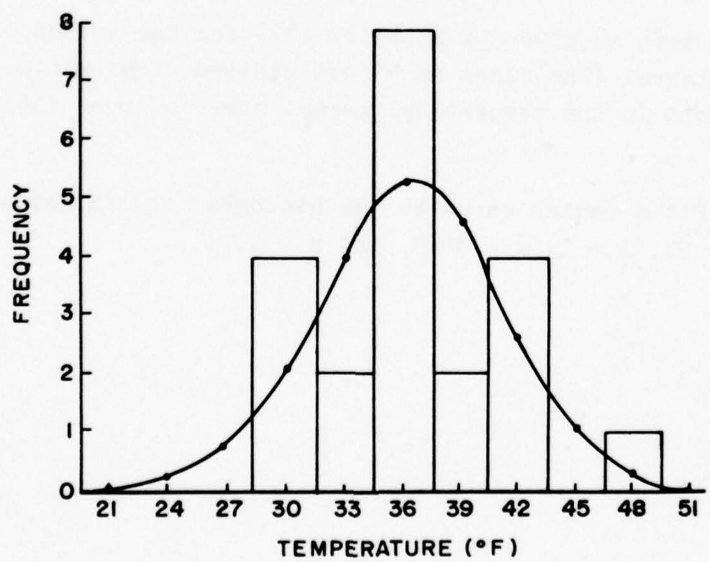


Figure 7. Normal Curve of Mean Temperatures, January 1943-1963, at Washington National Airport.

d. The cumulative normal distribution function gives the relative frequency (probability) that observations fall below (to the left of) any specified value. The cumulative normal distribution function, commonly called the normal distribution, is defined by:

$$(49) \qquad P = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2} dz$$

where, as before, $z = \frac{x - \bar{x}}{s}$. It must be remembered that Equation (49) is normalized, and consequently, the area from minus infinity to plus infinity is equal to one. This means that the area between minus infinity and any given z is equivalent to the relative frequency (probability) that the observation falls below the z value. Tables of normal distribution function are available in most statistical textbooks; however, it is convenient to use probability paper to determine the desired areas.

e. The cumulative normal distribution appears as a straight line on normal probability paper. The sample distribution can be plotted on probability paper and the degree to which the points form a straight line determines the fit of the sample to a normal distribution. The normalized Equation (49) is shown as a straight line on Figure 8, where areas below (to the left of) specified z values are given on the bottom scale. For example, the area below z = -2 is 0.023, below z = 0 is 0.50, z = 2 is 0.978, etc., for any z value.

<u>Example 29</u>: Can the temperature data in Table 5 be approximated by a normal distribution? Temperatures are ranked in increasing order of magnitude and cumulative probabilities are given by $P = m/1+N$, where m takes on values from 1 to N, as seen in Table 16. These temperatures were plotted on probability paper (see Figure 9) and the fit to a straight line is quite good, indicating the distribution of the data given in Table 5 is probably normal. The straight line shown on Figure 9 is for the cumulative normal distribution with the same mean and standard deviation as our sample.

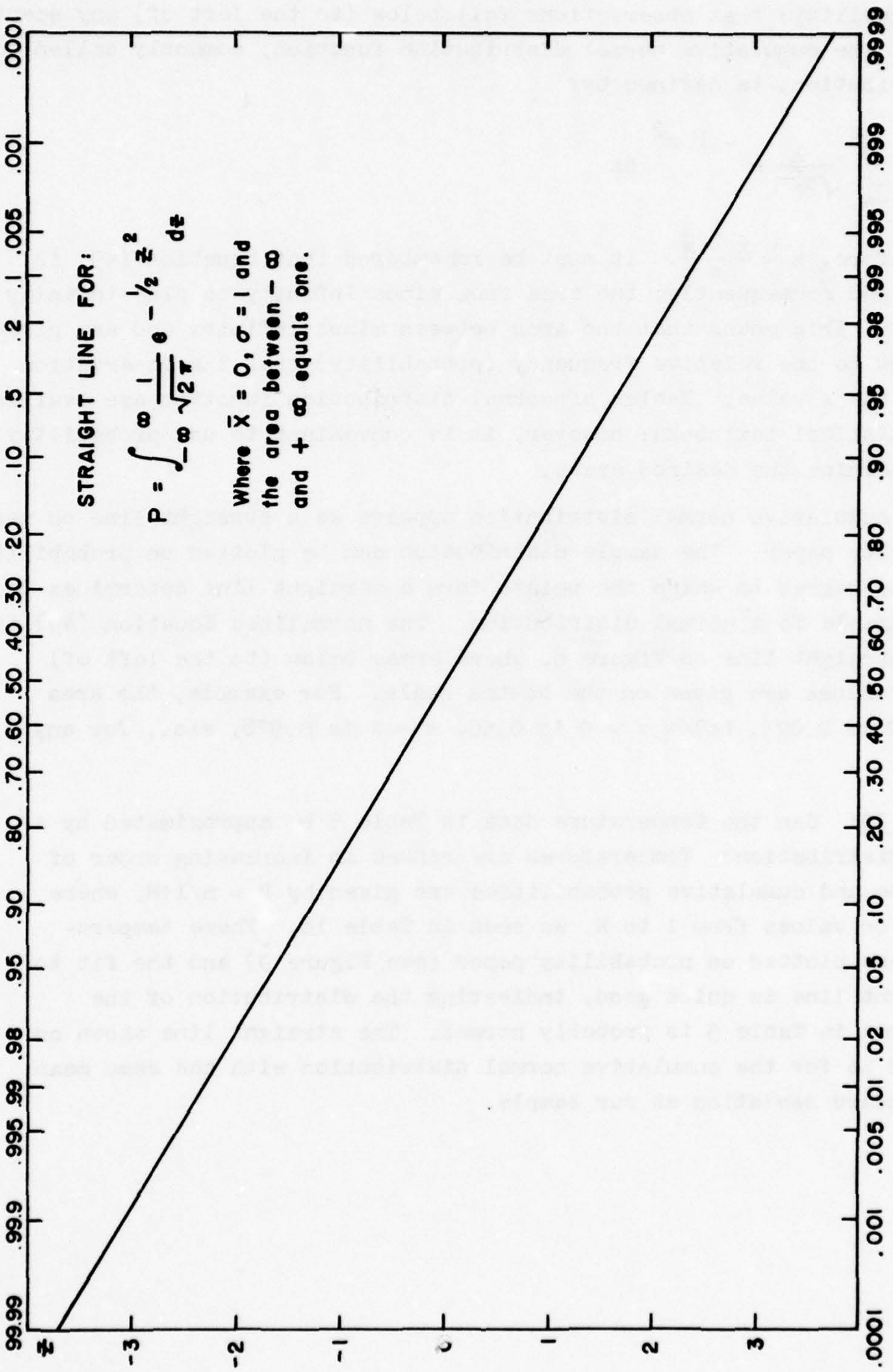


Figure 8. The Normalized Cumulative-Normal Distribution.

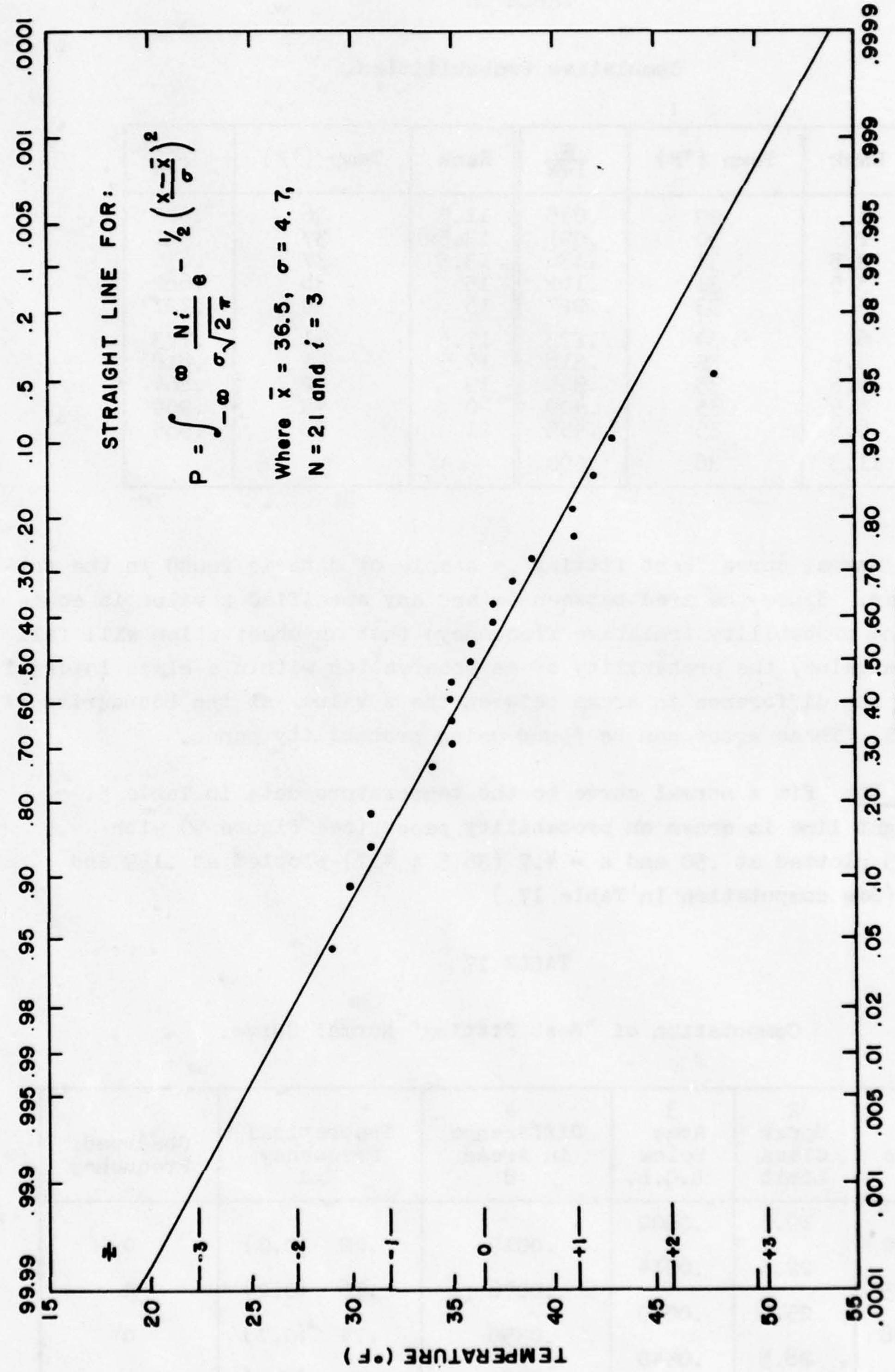


Figure 9. January Temperatures (°F), 1943-1963 at Washington National Airport.

TABLE 16

Cumulative Probabilities.

Rank	Temp (°F)	$\frac{m}{1+N}$	Rank	Temp (°F)	$\frac{m}{1+N}$
1	29	.045	11.5	36	.545
2	30	.091	13.5	37	.591
3.5	31	.136	13.5	37	.636
3.5	31	.182	15	38	.682
5	33	.227	16	39	.727
6	34	.273	17.5	41	.773
8.5	35	.318	17.5	41	.818
8.5	35	.364	19	42	.864
8.5	35	.409	20	43	.909
8.5	35	.455	21	48	.955
11.5	36	.500			

f. The normal curve "best fitting" a sample of data is found in the following manner. Since the area between $-\infty$ and any specified z value is equivalent to the probability (relative frequency) that an observation will fall below that z value, the probability of an observation within a class interval is equal to the difference in areas between the z values at the boundaries of the interval. These areas can be found using probability paper.

Example 30: Fit a normal curve to the temperature data in Table 5. A straight line is drawn on probability paper (see Figure 9) with $\overline{x}$ = 36.5 plotted at .50 and s = 4.7 (36.5 ± 4.7) plotted at .159 and .841. (See computation in Table 17.)

TABLE 17

Computation of "Best Fitting" Normal Curve.

1 Class	2 Upper Class Limit	3 Area Below L.C.L.	4 Difference in Areas. d	Theoretical Frequency Nxd	Observed Frequency
	19.5	.0002			
20-22			.0012	.02 (0.0)	0
	22.5	.0014			
23-25			.0076	.16 (0.2)	0
	25.5	.0090			
26-28			.0350	.74 (0.7)	0
	28.5	.0440			
29-31			.0960	2.02 (2.0)	4

TABLE 17 (Cont'd)

1 Class	2 Upper Class Limit	3 Area Below L.C.L.	4 Difference in Areas d	Theoretical Frequency Nxd	Observed Frequency
29-31			.0960	2.02 (2.0)	4
	31.5	.1400			
32-34			.1900	3.99 (4.0)	2
	34.5	.3300			
35-37			.2500	5.25 (5.3)	8
	37.5	.5800			
38-40			.2250	4.73 (4.7)	2
	40.5	.8050			
41-43			.1270	2.67 (2.7)	4
	43.5	.9230			
44-46			.0520	1.09 (1.1)	0
	46.5	.9840			
47-49			.0135	.28 (0.3)	1
	49.5	.9975			
50-52			.0022	.05 (0.0)	0
	52.5	.9997			

g. The Chi-square (χ^2) Test is used to test the "goodness of fit" of the observed frequencies (o) to the theoretical (expected) frequencies (e). The χ^2 statistic is given by:

$$(50) \qquad \chi^2 = \frac{(o_1-e_1)^2}{e_1} + \frac{(o_2-e_2)^2}{e_2} + \ldots + \frac{(o_k-e_k)^2}{e_k} = \sum_{j=1}^{k} \frac{(o_j-e_j)^2}{e_j}$$

where every e_j should be at least five, as required by the derivation of the chi-square distribution. The theoretical and observed frequencies agree exactly if $\chi^2 = 0$. If $\chi^2 > 0$, they do not agree, and the larger the value of χ^2, the greater the disagreement. In fact, when χ^2 is larger than certain limits, the null hypothesis (i.e., the observed frequencies do not differ significantly from the theoretical frequencies) is rejected, and the observed frequencies do differ significantly from the theoretical frequencies at the usually-accepted 5% (probability of 0.05) level. The distribution depends on the number of degrees of freedom (γ). The degree of freedom (γ) is equal to $k - 1 - m$, where k is the number of classes and m is the number of population statistics estimated from the sample. Tables of χ^2 for different degrees of freedom and various probability levels are included in most statistics books. A probability level of 0.05 means that the value of χ^2 for the given number of degrees of freedom will be equalled or exceeded, in random samples, only five times in 100.

Example 31: Can the observed frequencies in Example 30 be assumed to be from a normal distribution at the 5% level of significance? (It is necessary to combine some of the classes so that nearly five cases are in each group.)

TABLE 18

Computation of Chi-Square.

	Temperature (°F)					
	< 31	32-34	35-37	38-40	> 41	Total
Frequency Observed (o)	4.0	2.0	8.0	2.0	5.0	21
Theoretical (e)	2.9	4.0	5.3	4.7	4.1	21
Difference (d)	1.1	2.0	2.7	2.7	0.9	--
$\frac{(\text{Difference})^2}{\text{Theoretical}}$	$\frac{1.21}{2.9}$	$\frac{4.0}{4.0}$	$\frac{7.29}{5.3}$	$\frac{7.29}{4.7}$	$\frac{0.81}{4.1}$	--

$$\chi^2 = 0.417 + 1.0 + 1.375 + 1.551 + 0.198 = 4.54$$

$k = 5$ and $m = 2$ ($\overline{x}$ and s from sample used to compute theoretical frequencies) so $\gamma = 5 - 1 - 2 = 2$. From χ^2 tables, $\chi^2_{.05}$ for $\gamma = 2$ equals 5.99. Consequently, since $4.54 < 5.99$, we can assume that temperatures given in Table 5 came from a normal population.

7. Sampling.

a. In statistics, the sample estimate that is calculated usually possesses an appreciable error of estimate. Therefore, it is important to determine the most probable value of such statistical parameters as the mean and standard deviation. The mean for a series of values $X_1, X_2, \ldots, X_n$ is defined as

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \qquad \text{[See Equation (32)]}$$

and represents the best estimate of the mean of the population.

(1) The standard error of the population mean for independent data is usually given by:

(51) $$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{N}}$$

where σ_X is the standard deviation of the variable estimated by

$$s_X = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N} X_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} X_i\right)^2\right]} \qquad \text{[See Equation (40)]}$$

Because of the relative sparsity of most meteorological data, all consecutive values available usually are used to compute statistics such as the mean and standard deviation. In general, there may be persistence in the data and each value is correlated with adjacent values (autocorrelation). The autocorrelation, r_τ, at any time interval of lag τ can be estimated from the sample by:

$$r_\tau = \frac{1}{(N-\tau)\, s_X^2} \sum_{i=1}^{N-\tau} (X_i-\overline{X})(X_{i+\tau}-\overline{X}) \qquad \text{[See Equation (46)]}$$

and the standard error of the mean of the autocorrelated data becomes:

(52) $$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{N}}\sqrt{N + 2\sum_{\tau=1}^{N-1} (N-\tau)\, r_\tau}$$

Since persistence is probably the primary cause of the autocorrelation in most meteorological parameters, the autocorrelation at lag τ, r_τ, is related to that at lag one, r_1, by:

(53) $$r_\tau = r_1^{\tau}$$

where the autocorrelation at lag one is given by:

(54) $$r_1 = \frac{1}{(N-1)\, s_X^2} \sum_{i=1}^{N-1} (X_i-\overline{X})(X_{i+1}-\overline{X})$$

The standard error of the mean for autocorrelated data caused primarily by persistence is given by:

(55) $$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{N}}\sqrt{1 + \frac{2r_1}{1-r_1}\left[1 - \frac{1}{N}\left(\frac{1-r_1^{N}}{1-r_1}\right)\right]}$$

and, as given by Mitchell [37] for N > 10 and/or a small r_1, is roughly equivalent to:

(56) $$\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{N}} \sqrt{\frac{1 + r_1}{1 - r_1}}$$

(2) The 95% confidence limits for the mean of the population, as estimated by the sample mean, are given by:

(57) $$\overline{X}_{(0.95)} = \overline{X} \pm 1.96\ \sigma_{\overline{x}}$$

where $\sigma_{\overline{x}}$ is given by Equations (51), (52), (55), or (56), as appropriate. The 95% confidence interval is to be interpreted in the following manner. The interval, being a function of the random variable $\overline{x}$, is itself a random variable. If 100 samples were taken and 100 95%-confidence intervals were determined, then we would expect 95 of these intervals to include the population mean.

Example 32: What are the 95% confidence limits of the mean of the data given in Table 5? The mean monthly temperatures are independent; therefore, to determine the standard error of the mean of this data we use Equation (51). The mean is given by:

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i = \frac{766}{21} = 36.5$$

The standard deviation is given by:

$$s_x = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N} X_i^{\,2} - \frac{1}{N}\left(\sum_{i=1}^{N} X_i\right)^2\right]} = \sqrt{\frac{1}{20}\left[28{,}382 - \frac{1}{21}(586{,}756)\right]}$$

$$= \sqrt{\frac{441.24}{20}} = 4.7$$

The standard error of the mean as calculated from Equation (51) becomes:

$$\sigma_{\overline{x}} = \frac{4.7}{\sqrt{21}} = 1.03$$

The 95% confidence limits of the mean are given by Equation (57)

$$\overline{X} \pm 1.96\ \sigma_{\overline{x}} = 36.5 \pm 1.96\ (1.03) = 36.5 \pm 2.02 = 34.5 \text{ and } 38.5$$

In this case, the interval 34.5°F to 38.5°F has a probability of 0.95 of containing the population mean.

Example 33: What are the 95% confidence limits of the mean of the following mean temperatures for January 1962 at Washington National Airport (column X of Table 19)? It is assumed that persistence is the

TABLE 19

Computation of Variance, Lag r_1, and Standard Error.

Day	X	X^2	(1) $(X_i - \overline{X})$	(2) $(X_{i+1} - \overline{X})$	(1) X (2)
1	36	1296	(36-35)	(32-35)	-3
2	32	1024	(32-35)	(33-35)	+6
3	33	1089	(33-35)	(47-35)	-24
4	47	2209	(47-35)	(37-35)	+24
5	37	1369	(37-35)	(45-35)	+20
6	45	2025	(45-35)	(48-35)	+130
7	48	2304	(48-35)	(40-35)	+65
8	40	1600	(40-35)	(30-35)	-25
9	30	900	(30-35)	(19-35)	+80
10	19	361	(19-35)	(16-35)	+304
11	16	256	(16-35)	(24-35)	+209
12	24	576	(24-35)	(27-35)	+88
13	27	729	(27-35)	(32-35)	+24
14	32	1024	(32-35)	(46-35)	-33
15	46	2116	(46-35)	(35-35)	0
16	35	1225	(35-35)	(34-35)	0
17	34	1156	(34-35)	(29-35)	+6
18	29	841	(29-35)	(30-35)	+30
19	30	900	(30-35)	(33-35)	+10
20	33	1089	(33-35)	(35-35)	0
21	35	1225	(35-35)	(46-35)	0
22	46	2116	(46-35)	(42-35)	+77
23	42	1764	(42-35)	(35-35)	0
24	35	1225	(35-35)	(44-35)	0
25	44	1936	(44-35)	(42-35)	+63
26	42	1764	(42-35)	(44-35)	+63
27	44	1936	(44-35)	(33-35)	-18
28	33	1089	(33-35)	(30-35)	+10
29	30	900	(30-35)	(37-35)	-10
30	37	1369	(37-35)	(22-35)	-26
31	22	484	--	--	--
--	1083	39897	--	--	1070

primary cause of the autocorrelation, so Equation (56) is used to compute the standard error ($\sigma_{\overline{x}}$). The mean $\overline{X}$ is determined by

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{31} X_i = \frac{1083}{31} = 34.93 \approx 35$$

The variance, s_x^2, and the standard deviation, s_x, are found by using Equation (40).

$$s_x^{\,2} = \frac{1}{N-1}\left[\sum_{i=1}^{N} X_i^{\,2} - \frac{1}{N}\left(\sum_{i=1}^{N} X_i\right)^2\right] = \frac{1}{30}\left[39{,}897 - \frac{1}{31}(1083)^2\right]$$

$$= \frac{1}{30}\,[39{,}897 - 37{,}835] = \frac{2062}{30} = 68.7$$

$$s_x = \sqrt{68.7} = 8.3$$

The autocorrelation coefficient for a lag of one day, r_1, using Equation (46):

$$r_1 = \frac{1}{(N-1)\,s_x^{\,2}} \sum_{i=1}^{N-1} (X_i - \overline{X})(X_{i+1} - \overline{X}) = \frac{1}{(30)(68.7)}\,(1070)$$

$$= \frac{1070}{2061} = 0.519$$

The standard error of the mean, $\sigma_{\overline{x}}$, using Equation (56):

$$\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{N}}\sqrt{\frac{1 + r_1}{1 - r_1}} = \frac{8.3}{\sqrt{31}}\sqrt{\frac{1 + 0.519}{1 - 0.519}} = \frac{8.3}{\sqrt{31}}\sqrt{3.158} = 2.648$$

The 95% confidence limits of the mean are given by Equation (57):

$$\overline{X} \pm 1.96\,\sigma_{\overline{x}} = 35 \pm 1.96\,(2.648) = 35.00 \pm 5.19 = 29.81 \text{ and } 40.19$$

Therefore, the 95% confidence limits of the population mean for the autocorrelated data is 29.8°F and 40.2°F.

b. For an autocorrelated series of N terms the analyst determines the effective number, N', of independent terms. N', when used in Equation (51), gives the standard error of the mean of the autocorrelated series. Equating Equation (52) with Equation (51) gives

$$\text{(58)} \qquad N' = N\left[1 + \frac{2}{N}\sum_{\tau=1}^{N-1} (N - \tau)\, r_\tau\right]^{-1}$$

To find the standard error of the mean of persistent series, Equation (55) or (56) is used. Therefore, what number N', when substituted in Equation (51), will give a value for $\sigma_{\overline{x}}$ equal to the correct value given by Equation (55) or (56)? Equating Equation (55) with (51) gives

$$\text{(59)} \qquad N' = N\left[1 + \frac{2\,r_1}{1-r_1}\left(1 - \frac{1}{N}\,\frac{1-r_1^{\,N}}{1-r_1}\right)\right]^{-1}$$

and for $N > 10$ and/or a small r_1, we find from Equation (59)

$$N' = N\left(\frac{1-r_1}{1+r_1}\right) \quad (60)$$

Example 34: What are the effective number of independent daily temperatures in the data series given in Example 33? Since $N > 10$, it is appropriate to use Equation (60).

$$N' = N\left(\frac{1 - r_1}{1 + r_1}\right) = 31\left(\frac{1 - 0.519}{1 + 0.519}\right) = 9.8 \approx 10$$

Therefore, the temperatures on every third day are independent of each other.

c. After calculating a sample standard deviation, s_x, the analyst may wish to determine if its value is significantly different from the population value, σ_x. The confidence bands of the standard deviation can be found by means of the chi-square, χ^2, test. The test consists of verifying whether

$$\sqrt{\frac{Ns_x^2}{\chi_1^2}} > \sigma_x > \sqrt{\frac{Ns_x^2}{\chi_2^2}} \quad (61)$$

where χ_1^2 and χ_2^2 are the lower and upper limits, respectively, of the chi-square distribution (N - 1 degrees of freedom) appropriate to a desired confidence level, e.g., if s_x is computed from an autocorrelated series then the test must be modified by replacing N by N' from Equations (58), (59), or (60) and the degrees of freedom taken as N' - 1.

Example 35: Find the 95% confidence limits of the standard deviation for the temperature data in Table 5. Since the data are independent, it is appropriate to use $N = 31$ in Equation (61) and $\chi_{2.5}^2$ and $\chi_{97.5}^2$ for $N - 1 = 30$ degrees of freedom with $s_x^2 = 22.1$ to determine the 95% confidence level.

$$\sqrt{\frac{(31)(22.1)}{16.79}} > \sigma_x > \sqrt{\frac{(31)(22.1)}{46.98}}$$

$$\sqrt{40.8} > \sigma_x > \sqrt{14.6}$$

$$6.4 > \sigma_x > 3.8 \quad \text{for } s_x = 4.7$$

The 95% confidence limits are 3.8°F and 6.4°F.

Example 36: Find the 95% confidence limits of the standard deviation

of the autocorrelated data given in Example 33. Since these data are autocorrelated, it is appropriate to use N' = 10, as found in Example 34 by Equation (60), and $\chi_{2.5}^2$ and $\chi_{97.5}^2$ for N' - 1 = 9 degrees of freedom with $s_x^2 = 68.7$ to find the 95% confidence level.

$$\sqrt{\frac{(10)(68.7)}{2.70}} > \sigma_x > \sqrt{\frac{(10)(68.7)}{19.02}}$$

$$\sqrt{254.4} > \sigma_x > \sqrt{35.8}$$

$$15.95 > \sigma_x > 5.98$$

Therefore, the interval 6°F to 16°F has a probability of 0.95 of containing the population standard deviation.

8. Statistical Inference.

a. Since dealing with populations themselves is generally impractical in statistics, absolute proofs of statements about populations are rarely possible. Therefore, an investigator must use statistical tests to estimate the accuracy of his statements. One such test is the use of statistical inference.

b. Consider any population. Suppose that a value, "h," has been used for a parameter of a population and that an analyst wants to test this value. He has two ways of making a wrong decision. He may reject "h" when it is right (called the Type I or α error), or he may accept "h" when it is wrong (called the Type II or β error). Statistical inference enables him to limit α, the chance of rejecting a correct hypothesis, to any percent he chooses. Once α is chosen, he can also determine β, the chance of accepting an incorrect hypothesis. For any particular α, the value of β will depend on how wrong the hypothesis is. Note that this test does not tell him the chance of the hypothesis being right or wrong! It tells him only the chance of rejecting it when it is right, or accepting it when it is wrong. In general, as α is decreased, β will be increased. The analyst usually decides how high he wants α to be and sets up his test based on this decision. If he wants α to be 5%, he is said to have set up a level of significance of 5%.

c. To apply the test, we take a sample and compute from it the parameter in question and the standard error, σ_1, of this parameter. Assume the sample distribution of most parameters to be normal or nearly so. Using the hypothesis "h" as the mean of the parameter in question and the standard error of this parameter [paragraph 7a(1) of this chapter], we draw the sample distribution as a straight line on normal probability paper. A 5% level of significance means that the hypothesis will be accepted if the sample value of the parameter falls within the 2.5 through 97.5% ("h" ± 1.96 σ_1) limits on our

distribution. In other words, _if_ our hypothesis is correct, the analyst will accept it 95% of the time because 95% of the sample values will fall within the limits. Suppose, however, that the hypothesis is wrong, and the value of the population parameter is not "h" but "m." The sample distribution of the parameter is still normal and has the same standard error, but it now has a mean of "m." The probability of occurrence of "h ± 1.96 σ_1" on our new distribution is β.

Example 37: Define a population as the monthly temperatures for all years at Washington National Airport. Suppose that 38°F has been used as the mean of these mean monthly temperatures. Test this value to decide whether or not to accept it. For this example, set up a level of significance of 5% so that if the 38°F is correct, we will accept it 95% of the time.

(a) Use the sample years 1943-1963 to decide if the hypothesis should be accepted or rejected. The mean ($\bar{x}$) of this sample is computed to be 36.5°F and s, its standard deviation, [illegible] The standard error of the mean of the population is given by

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \qquad \text{[See Equation (51)]}$$

where $\sigma_{\bar{x}}$ is the standard error of the mean, N is the number in the sample, and σ (the standard deviation of the population) is approximated by s. $\sigma_{\bar{x}}$ is computed to be 1.03 (see Example 32). Using as the mean, the hypothesis 38°, and $\sigma_{\bar{x}} = 1.03$, draw the sample distribution as a straight line on normal probability paper (Line "A" on Figure 10). Note that a level of significance of 5% represents a value of 36.0 (h - 1.96 $\sigma_{\bar{x}}$) at the 2.5% limit and a value of 40.0 (h + 1.96 $\sigma_{\bar{x}}$) at the 97.5% limit. Any sample value between these limits causes the analyst to accept the hypothesis. Since the sample value is 36.5°, he would accept 38°.

(b) What is the chance of his accepting this hypothesis if it is 1, 2, or 3° off? Using the same $\sigma_{\bar{x}}$, but this time 39° as the mean, draw another sample distribution (Line "B" on Figure 10). The percentage of time that the values 36.0° through 40.0° occur on this distribution is the β error. In this case, the β error is 83% minus ≈ 0.2% or 82.8%. In other words, the analyst runs an 82.8% chance of accepting 38°, even if the actual mean is one degree off, i.e., either 37° or 39°. Similarly, β for two degrees off would be 50% (Line "C" on Figure 10), and for three degrees off would be 16% (Line "D" on Figure 10).

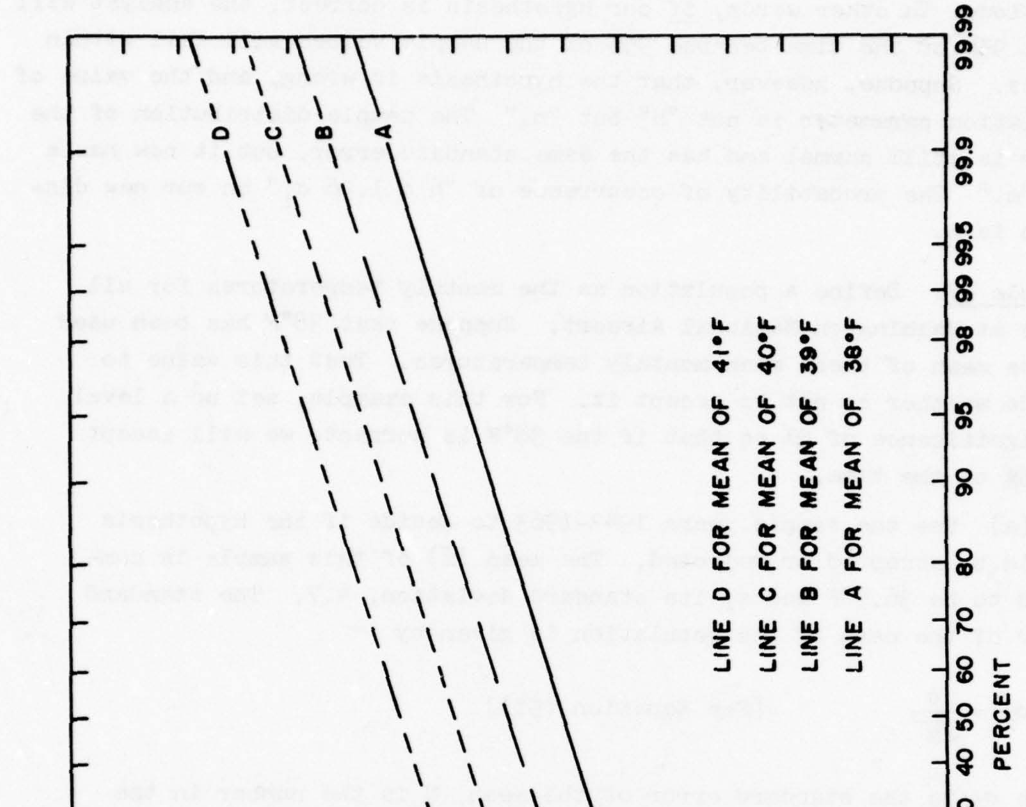


Figure 10. Sample Distribution of the Mean of the Mean Monthly Temperatures for Washington National Airport.

9. Regression and Correlation.

a. In meteorology and climatology it is often necessary to determine if two or more weather factors or parameters are related and, if so, how they are related. A relationship between two or more parameters is called correlation. An equation that can be used to predict one parameter from information about the other and to estimate the range of the error in the prediction can be derived from this correlation.

b. If data for two variables are plotted on a scatter diagram (see Figure 11) and appear to fall along a straight line, a linear equation probably gives the best relationship. A line of regression is the line of best fit of the data and is given by the slope-intercept equation of the straight line,

$$Y = m X + b \tag{62}$$

where m is the slope with respect to the X axis and b is the Y-intercept. Usually the equation of the line is found by the method of least squares, which gives a line such that the sum of the squares of the deviations of the observed values from the line is a minimum. The constants, m and b, are determined by solving simultaneously the following two normal equations of the least square line:

$$\Sigma Y = m \Sigma X + b N \tag{63}$$

$$\Sigma X Y = m \Sigma X^2 + b \Sigma X \tag{64}$$

$$b = \frac{\Sigma Y \; \Sigma X^2 - \Sigma X \; \Sigma XY}{N \Sigma X^2 - (\Sigma X)^2} \tag{65}$$

$$m = \frac{\Sigma Y - Nb}{\Sigma X} \tag{66}$$

Equations (65) and (66), when substituted into the slope-intercept Equation (62), give the line of best fit for the regression of Y on X by the least square method.

c. The standard error of estimate, which is analogous to the standard deviation, is a measure of the scatter about the line of regression of Y on X and is given by

$$s_{Y \cdot X} = \sqrt{\frac{\Sigma Y^2 - b \Sigma Y - m \Sigma X Y}{N}} \tag{67}$$

If lines are drawn parallel to the regression line of Y or X at distances equal to $s_{Y \cdot X}$ above and below the line measured in the Y direction, about 68% of the data points fall between the two lines (Figure 11).

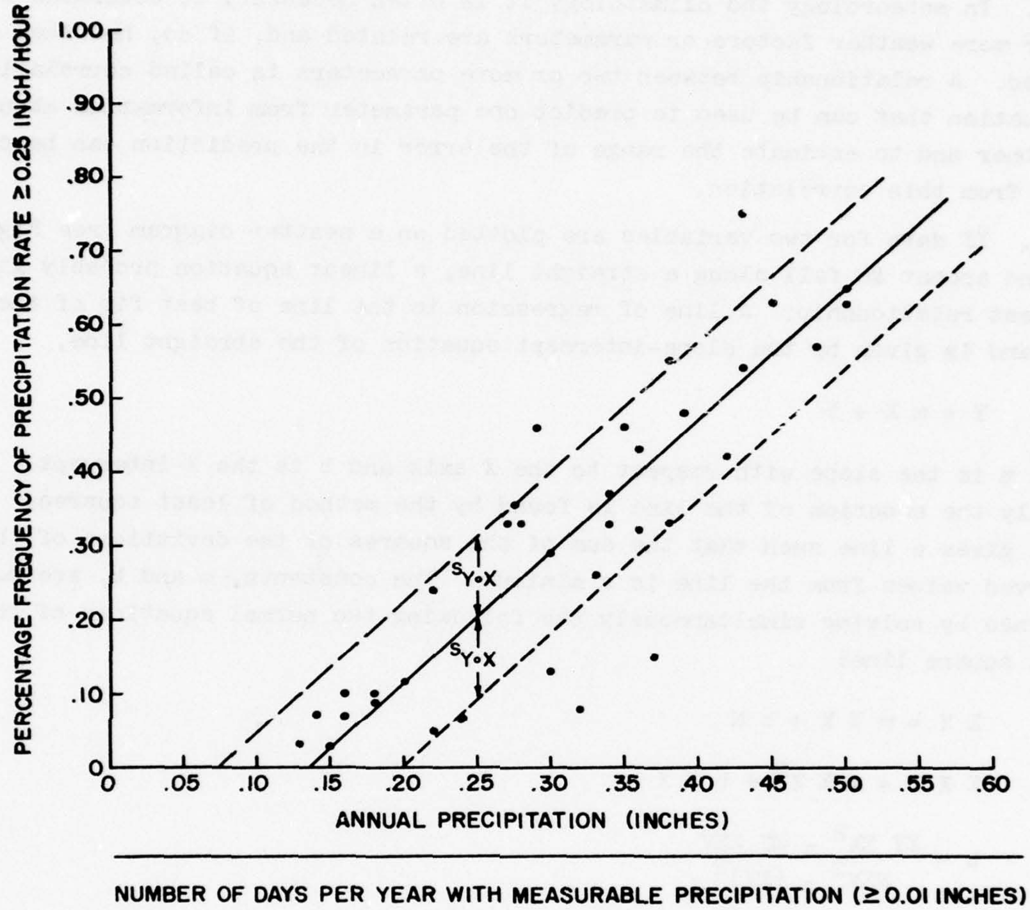


Figure 11. Line of Regression.

d. The linear correlation coefficient measures the relationship between two factors or parameters. It is dimensionless, varies between -1 and +1, and is given by

$$(68) \qquad r = \sqrt{1 - \frac{s_{Y \cdot X}^{\,2}}{s_Y^{\,2}}}$$

where s_Y is the standard deviation of the parameter Y. It must be remembered that r is a linear correlation coefficient, and a value near zero does not always mean no correlation because there may be a high nonlinear correlation

TABLE 20

Computation of Coefficients, Standard Error of Estimate, and Correlation Coefficient.

	X	Y	XY	X^2	Y^2
1	0.33	0.26	.0858	.1089	.0676
2	0.16	0.07	.0112	.0256	.0049
3	0.43	0.54	.2322	.1849	.2916
4	0.42	0.42	.1764	.1764	.1764
5	0.13	0.03	.0039	.0169	.0009
6	0.22	0.24	.0528	.0484	.0576
7	0.43	0.75	.3225	.1849	.5625
8	0.27	0.33	.0891	.0729	.1089
9	0.28	0.33	.0924	.0784	.1089
10	0.18	0.10	.0180	.0324	.0100
11	0.18	0.09	.0162	.0324	.0081
12	0.38	0.33	.1254	.1444	.1089
13	0.14	0.07	.0098	.0196	.0049
14	0.30	0.29	.0870	.0900	.0841
15	0.50	0.63	.3150	.2500	.3939
16	0.37	0.15	.0555	.1369	.0225
17	0.46	0.77	.3542	.2116	.5929
18	0.48	0.57	.2736	.2304	.3249
19	0.35	0.46	.1610	.1225	.2116
20	0.38	0.55	.2090	.1444	.3025
21	0.39	0.48	.1872	.1521	.2304
22	0.29	0.46	.1334	.0841	.2116
23	0.22	0.05	.0110	.0484	.0025
24	0.32	0.08	.0256	.1024	.0064
25	0.16	0.10	.0160	.0256	.0100
26	0.36	0.32	.1152	.1296	.1024
27	0.30	0.13	.0390	.0900	.0169
28	0.24	0.07	.0168	.0576	.0049
29	0.45	0.63	.2835	.2025	.3969
30	0.15	0.03	.0045	.0225	.0009
31	0.34	0.33	.1122	.1156	.1089
32	0.36	0.43	.1548	.1296	.1849
33	0.34	0.37	.1258	.1156	.1369
--	10.310	10.460	3.9160	3.5875	4.8602

between the parameters.

Example 38: Is there a relationship between the percentage frequency of precipitation rate ≥ 0.25 inch per hour and the ratio of annual precipitation (inches) and days per year with measurable precipitation (≥ 0.01 inch)? Data were compiled for 33 U.S. Weather Bureau first-order weather stations (Table 20) with:

$$X = \frac{\text{Annual precipitation (inches)}}{\text{Number of days with measurable precipitation}}$$

Y = Percentage frequency of precipitation rate ≥ 0.25 inch per hour

Values for X and Y are plotted on a scatter diagram (Figure 11) and the points form a straight line.

$$N = 33,\ s_Y^{\ 2} = \frac{\Sigma Y^2 - \frac{1}{N}(\Sigma Y)^2}{N - 1} = \frac{4.8602 - \frac{1}{33}(10.46)^2}{32} = 0.048276$$

$$b = \frac{\Sigma Y\ \Sigma X^2 - \Sigma X\ \Sigma XY}{N\Sigma X^2 - (\Sigma X)^2} = \frac{(10.46)(3.5875) - (10.31)(3.9160)}{(33)(3.5875) - (10.31)^2} = -0.235598$$

$$m = \frac{\Sigma Y - Nb}{\Sigma X} = \frac{10.46 - (33)(-0.235598)}{10.31} = 1.7686$$

$$Y = m\,X + b$$

<u>Line of Regression</u>

$$Y = 1.769\ X - 0.2356$$

<u>Standard Error of Estimate</u>

$$s_{Y\circ X} = \sqrt{\frac{\Sigma Y^2 - b\Sigma Y - m\Sigma XY}{N}}$$

$$= \sqrt{\frac{4.8602 - (-0.235598)(10.46) - (1.7686)(3.916)}{33}} = 0.109895$$

<u>Linear Correlation Coefficient</u>

$$r = \sqrt{1 - \frac{s_{Y\circ X}^{\ 2}}{s_Y^{\ 2}}} = \sqrt{1 - \frac{0.0120769}{0.048276}} = 0.87$$

Therefore, there appears to be a good relation between the ratio and the percentage frequency of a precipitation rate of ≥ 0.25 inches per hour.

<u>Example 39</u>: What is the relationship between the ratio given in Example 38 (annual precipitation in inches divided by number of days with measurable precipitation) when: (a) the percentage frequency of precipitation rate ≥ 0.50 inches per hour, and (b) the percentage frequency of precipitation rate ≥ 1.00 inches per hour. The computations yield the following:

First, ratio versus percentage frequency of precipitation rate ≥ 0.50 inches per hour.

<u>Line of Regression</u>
$Y = 0.763\ X - 0.122$

<u>Standard Error of Estimate</u>
$s_{Y \circ X} = 0.0568$

<u>Linear Correlation Coefficient</u>
$r = 0.82$

Second, ratio versus percentage frequency of precipitation rate ≥ 1.00 inches per hour.

<u>Line of Regression</u>
$Y = 0.223\ X - 0.0388$

<u>Standard Error of Estimate</u>
$s_{Y \circ X} = 0.0230$

<u>Linear Correlation Coefficient</u>
$r = 0.72$

e. Some nonlinear equations can be reduced to linear form. After the data for two variables are plotted, it may be apparent that a straight line does not fit the data, but a curve might. The data may then be plotted on log-log graph paper or semi-log graph paper.

(1) If a straight line results on log-log paper the equation

(69) $$Y = uX^{v}$$

satisfies the data. Equation (69) can be reduced by taking logarithms of both sides of the equation and the resulting equation becomes:

(70) $$\log Y = \log u + v \log X$$

Setting $\log Y = Y'$, $\log u = u'$, and $\log X = X'$, the resulting linear equation becomes

(71) $$Y' = u' + vX'$$

As seen by Equation (70), the logarithms of Y and X are used to calculate the slope (v) and the Y-intercept (log u) of Equation (71). After the line of best fit Equation (71) has been derived, it is transformed back to Equation (69) form.

(2) If a straight line on log-log paper is not obtained, the data can be plotted on semi-log paper. First, with X on the linear scale and Y on the log scale. If the result is not a straight line, reverse the procedure. If a straight line results from either of the two graphs, then Equation

(72) $Y = uv^X$ (X on the linear scale)

or

(73) $X = uv^Y$ (Y on the linear scale)

will satisfy the data. Equations (72) and (73) can be reduced by taking logarithms of both sides.

(74) $\log Y = \log u + X \log v$

Setting $\log Y = Y'$, $\log u = u'$, and $\log v = v'$, the resulting linear equation becomes

(75) $Y' = u' + v' X$

Logarithms are used for the dependent variable and the actual value for the independent variable to calculate the slope (log v) and the Y-intercept (log u). The line of regression, Equation (75), is then derived and transformed to the exponential form, Equation (72) or (73), whichever is appropriate.

f. The calculated coefficient (r) is only a sample statistic. The population correlation coefficient (ρ) is needed, but is not known. However, confidence limits can be calculated by using the Fisher's Z transformation [11].

(76) $Z = \frac{1}{2} [\ell n\,(1 + r) - \ell n\,(1 - r)]$

Values for Z are given in Table 21. Equation (76) is approximately normally distributed with mean

(77) $u_Z = \frac{1}{2} [\ell n\,(1 + \rho) - \ell n\,(1 - \rho)]$

and standard deviation

(78) $\sigma_Z = \dfrac{1}{\sqrt{N - 3}}$

<u>Example 40</u>: What are the confidence limits for the correlation coefficient of 0.87 with N = 33 given in our computation example? The 95% confidence limits for u_Z are given by

$$Z \pm 1.96\,\sigma_Z = \frac{1}{2} [\ell n\,(1 + 0.87) - \ell n\,(1 - 0.87)] \pm 1.96 \frac{1}{\sqrt{30}}$$

$$= 1.333 \pm 1.96 \times \frac{1}{5.48} = 0.973 \quad \text{and} \quad 1.693$$

If

$$u_Z = \tfrac{1}{2}\,[\ell n\,(1+\rho) - \ell n\,(1-\rho)] = 0.973, \quad \rho = 0.75$$

and if,

$$u_Z = \tfrac{1}{2}\,[\ell n\,(1+\rho) - \ell n\,(1-\rho)] = 1.693, \quad \rho = 0.93$$

Therefore, the 95% confidence limits for the population correlation coefficient ρ are 0.73 and 0.93 corresponding to the sample correlation coefficient of 0.87.

TABLE 21

Fisher's Z Transformation.

r	0.00000	0.01000	0.02000	0.03000	0.04000
0.00000	0.00000	0.01000	0.02000	0.03001	0.04002
0.10000	0.10034	0.11045	0.12058	0.13074	0.14093
0.20000	0.20273	0.21317	0.22366	0.23419	0.24477
0.30000	0.30952	0.32055	0.33165	0.34283	0.35409
0.40000	0.42365	0.43561	0.44769	0.45990	0.47223
0.50000	0.54931	0.56273	0.57634	0.59015	0.60416
0.60000	0.69315	0.70892	0.72501	0.74142	0.75817
0.70000	0.86730	0.88718	0.90764	0.92873	0.95048
0.80000	1.09861	1.12703	1.15682	1.18814	1.22117
0.90000	1.47222	1.52752	1.58903	1.65839	1.73805

r	0.05000	0.06000	0.07000	0.08000	0.09000
0.00000	0.05004	0.06007	0.07011	0.08017	0.09024
0.10000	0.15114	0.16139	0.17167	0.18198	0.19234
0.20000	0.25541	0.26611	0.27686	0.28768	0.29857
0.30000	0.36544	0.37689	0.38842	0.40006	0.41180
0.40000	0.48470	0.49731	0.51007	0.52298	0.53606
0.50000	0.61838	0.63283	0.64752	0.66246	0.67767
0.60000	0.77530	0.79281	0.81074	0.82911	0.84796
0.70000	0.97296	0.99622	1.02033	1.04537	1.07143
0.80000	1.25615	1.29334	1.33308	1.37577	1.42193
0.90000	1.83178	1.94591	2.09230	2.29756	2.64665

NOTE: Z is negative when r is negative.

10. Time Series.

a. Periodicity of heavenly bodies is common knowledge. Night follows day and seasons change regularly. These periodicities are reflected in surface temperature changes that do not follow an exact period; yet, the probability of certain temperature changes tend to be periodic. Harmonic Analysis and Spectrum Analysis isolate those portions of the march of a meteorological

variable such as temperature, ceiling height, and wind speeds, to name only a few. While only a brief discussion of harmonic and spectrum analysis is included in this pamphlet, those readers desiring further information concerning these subjects are referred to the complete discussion contained in "Handbook of Statistical Methods in Meteorology" [11] and "Some Applications of Statistics to Meteorology" [38].

b. Harmonic Analysis generates a series of cosine functions that, when summed together, equal the original data of the series. A periodicity may be simple, composed of a single cosine curve, or may be more complex, containing many harmonics. According to mathematical principles, any function that is defined at every point in the interval can be represented by an infinite series of sine and cosine functions. A Fourier Series, established by the Fourier Analysis, represents this function. In the case of meteorological data, only a finite number of discrete points exists, so a finite number of harmonics (half the number of observations) will account for all the variation. For practical purposes, however, the first two or three harmonics account for most of the variation, and physical meaning can be generally attributed to these harmonics. Higher harmonics are reflections of "noise" in the observations or are purely of mathematical importance. The value X at time t_j equals the mean $\overline{X}$ plus the sum of all N/2 harmonics, thus:

$$(79) \qquad X_j = \overline{X} + \sum_{i=1}^{N/2} \left[A_i \text{ Sin } \left(\frac{2\pi}{P} i \, t_j\right) + B_i \text{ Cos } \left(\frac{2\pi}{P} i \, t_j\right)\right]$$

where P is the fundamental or total period of the function (the total length of record and not necessarily equal to N). For simplicity of writing, this series can be written as:

$$(80) \qquad X_j = \overline{X} + \sum_{i=1}^{N/2} \left\{ C_i \text{ Cos } \left[\frac{2\pi}{P} i \, (t_j - t_{max_i})\right]\right\}$$

where

$$(81) \qquad C_i = \sqrt{(A_i^2 + B_i^2)}$$

and:

$$(82) \qquad t_{max_i} = \frac{P}{2\pi i} \text{ ARCSIN } (A_i/C_i) = \frac{P}{2\pi i} \text{ ARCTAN } (A_i/B_i)$$

where the trigonometric functions give the same angular value. (On some occasions, the angles are double-valued.) This comparison assures the proper

quadrant of the determined angle, which is the time of the maximum for that particular harmonic. The problem is to compute A_i and B_i from which all information can be obtained. These are calculated from:

$$A_i = \frac{2}{N} \sum_{j=1}^{N} \left[X_j \operatorname{Sin} \left(\frac{2\pi}{P}\, i\, t_j \right) \right] \tag{83}$$

$$B_i = \frac{2}{N} \sum_{j=1}^{N} \left[X_j \operatorname{Cos} \left(\frac{2\pi}{P}\, i\, t_j \right) \right] \tag{84}$$

The origin of time is immaterial as the final harmonic series will always sum to the original data with changes to A_i, B_i, and t_{max_i}. How many harmonics are needed? This can best be answered by calculating the variance of each harmonic (each of which is independent) and forming a ratio of the variance explained by that harmonic to the total variance of X. While all harmonics explain 100% of the variance, the first three harmonics usually account for 90% or more. The variance for each, except the N/2 harmonic, is $C_i^2/2$. For the N/2 harmonic, the variance is C_i^2. Thus, the percentage variance of the i^{th} harmonic is:

$$100 \times \frac{C_i^2 \,(N - 1)}{2\left[\sum_{j=1}^{N} X_j^2 - \frac{1}{N} \left(\sum_{j=1}^{N} X_j \right)^2 \right]} \tag{85}$$

Example 41: Consider the mean number of January days that paradrop criteria are met for each hour at Seymour-Johnson AFB. The mean number of days at each hour is influenced by the three limiting meteorological parameters in the paradrop criteria — ceiling ≥ 2000 feet, visibility ≥ 3 miles, and surface winds < 10 knots. Mean number of days of paradrop criteria for January at each hour (local time) at Seymour-Johnson AFB are given in Table 22. In this example, P = 24 since the time units are 'hours' and the total time is one day (24 hours). By coincidence, there is an observation for each hour so here we find that N = 24. To generate the first harmonic, we have:

$$A_1 = \frac{2}{24} \sum_{j=1}^{24} \left[X_j \operatorname{Sin} \left(\frac{2\pi}{24}\, t_j \right) \right]$$

TABLE 22

January Paradrop Criteria.

Hour "t"	Mean No. of Days "X"	Hour "t"	Mean No. of Days "X"
Midnight	21.6	Noon	15.6
1	21.1	13	15.7
2	21.2	14	16.2
3	20.8	15	16.5
4	20.3	16	18.3
5	20.4	17	20.5
6	20.0	18	23.0
7	19.2	19	23.1
8	19.5	20	23.4
9	18.0	21	22.4
10	17.4	22	22.4
11	17.5	23	21.5

$$B_1 = \frac{2}{24} \sum_{j=1}^{24} \left[X_j \operatorname{Cos}\left(\frac{2\pi}{24} t_j\right)\right]$$

Considering only a few sine terms, we have:

$$\text{Term 1} = 21.6 \times \operatorname{Sin}\left[\frac{\pi}{12}(0)\right] = 0.00$$

$$\text{Term 2} = 21.1 \times \operatorname{Sin}\left[\frac{\pi}{12}(1)\right] = 5.45$$

$$\text{Term 3} = 21.2 \times \operatorname{Sin}\left[\frac{\pi}{12}(2)\right] = 10.55$$

$$\vdots \qquad\qquad \vdots$$

$$\text{Term 24} = 21.5 \times \operatorname{Sin}\left[\frac{\pi}{12}(23)\right] = -5.56$$

After obtaining every A_1, B_1, and t_1, the remainder of the calculations can be completed:

$C_1 = 2.901$ days	$T_{max_1} = -.910$ hours	% $vrnce_1 = 74.6$
$C_2 = 1.492$ days	$T_{max_2} = 7.223$ hours	% $vrnce_2 = 19.7$
$C_3 = 0.589$ days	$T_{max_3} = 2.383$ hours	% $vrnce_3 = 3.1$

The first three harmonics account for 97.4% of the variance. The actual times of the maximum for the various harmonics are calculated from midnight, the origin of the time for this example.

$$T_{max_1} = 2305$$
$$T_{max_2} = 0713$$
$$T_{max_3} = 0223$$

A graph of the mean number of days (observed), along with a sketch of three harmonics, is shown in Figure 12.

c. Spectrum analysis indicates whether a harmonic is important or unimportant. The graph of the percent variance of each harmonic to the harmonic number is the spectrum of that data. The spectrum of a time series shows the contributions of the various frequencies. Figure 13 presents the spectra for all 12 harmonics of the time series of the data in Example 41. Only the first three harmonics have a significant percent of the variance associated with them. From the 4th and higher harmonics, the variance is more likely a reflection of noise in the data. The actual determination of spectrum analysis is a harmonic analysis of the autocorrelation resulting in a smoothing of the calculated variances. The procedure for calculating the spectrum analysis is quite straightforward. First, lags from 0 to m (where m = P/2) are determined. Since the autocorrelation r_τ between hour/A and hour/B is exactly the negative of the autocorrelation between hour/B and hour/A, a symmetrical function results with the coefficients of all sine terms being zero (0). Thus, where 2m is the fundamental period and r_L is the autocorrelation of the coefficient of lag L, the actual calculation becomes:

(86) $$B_i = \frac{r_0}{m} + \frac{r_m}{m} (-1)^i + \frac{2}{m} \sum_{L=1}^{m-1} \left[r_L \operatorname{Cos} \left(\frac{\pi}{m} i L \right) \right]$$

In the case of B_0 and B_m the coefficients resulting from the formula must be divided by two. If B_i is plotted as a function of i, the resulting curve is a smoothed version of the original series and is scaled (or normalized). A better estimate of the smoothed spectrum function is obtained by weighting the coefficients:

(87) $$W_i = \tfrac{1}{4} B_{i-1} + \tfrac{1}{2} B_i + \tfrac{1}{4} B_{i+1}$$

<u>NOTE</u>: The W_i's are superimposed over the spectrum of the variances in Figure 13.

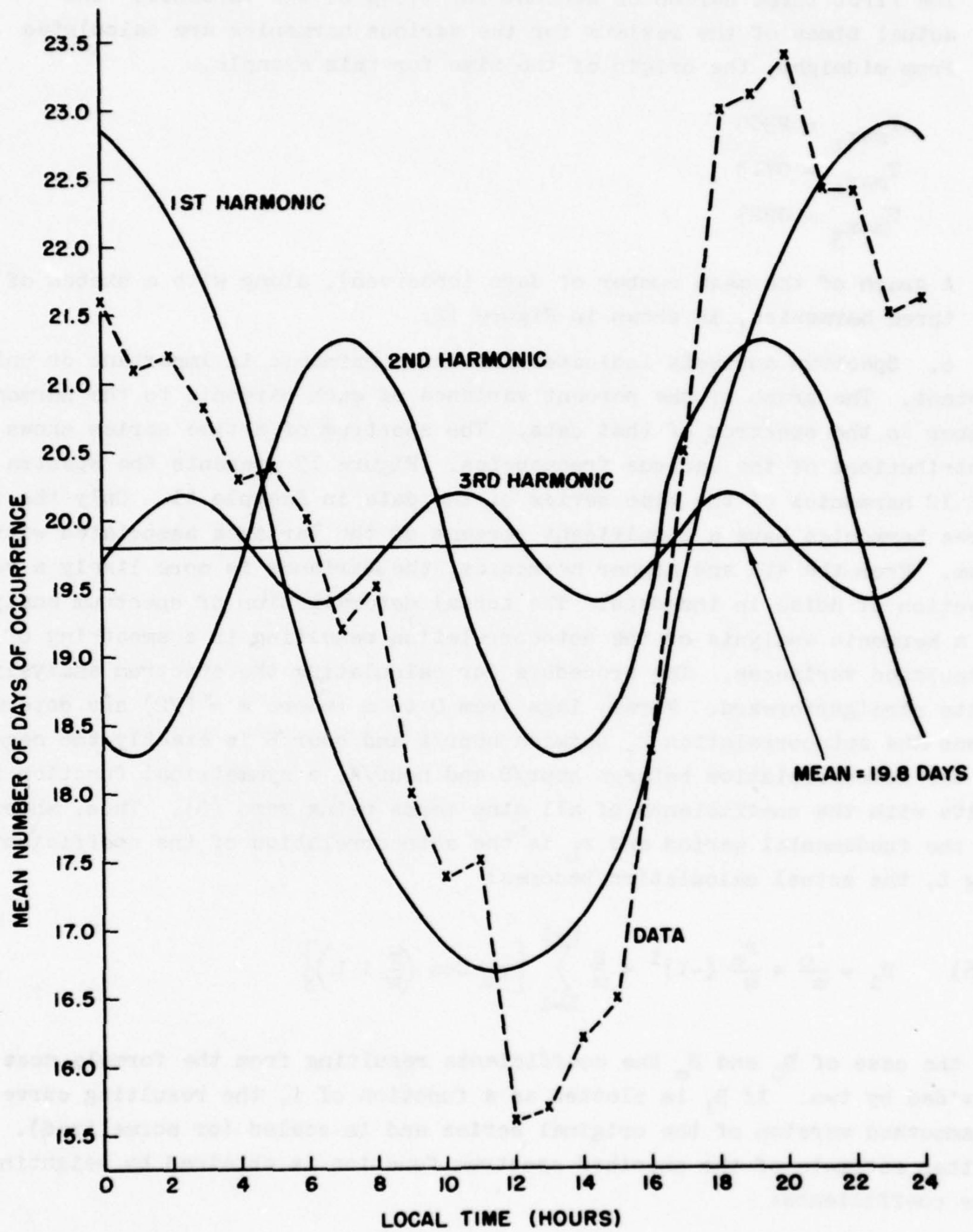


Figure 12. Mean Number of Days of Paradrop Criteria at Seymour-Johnson AFB, January.

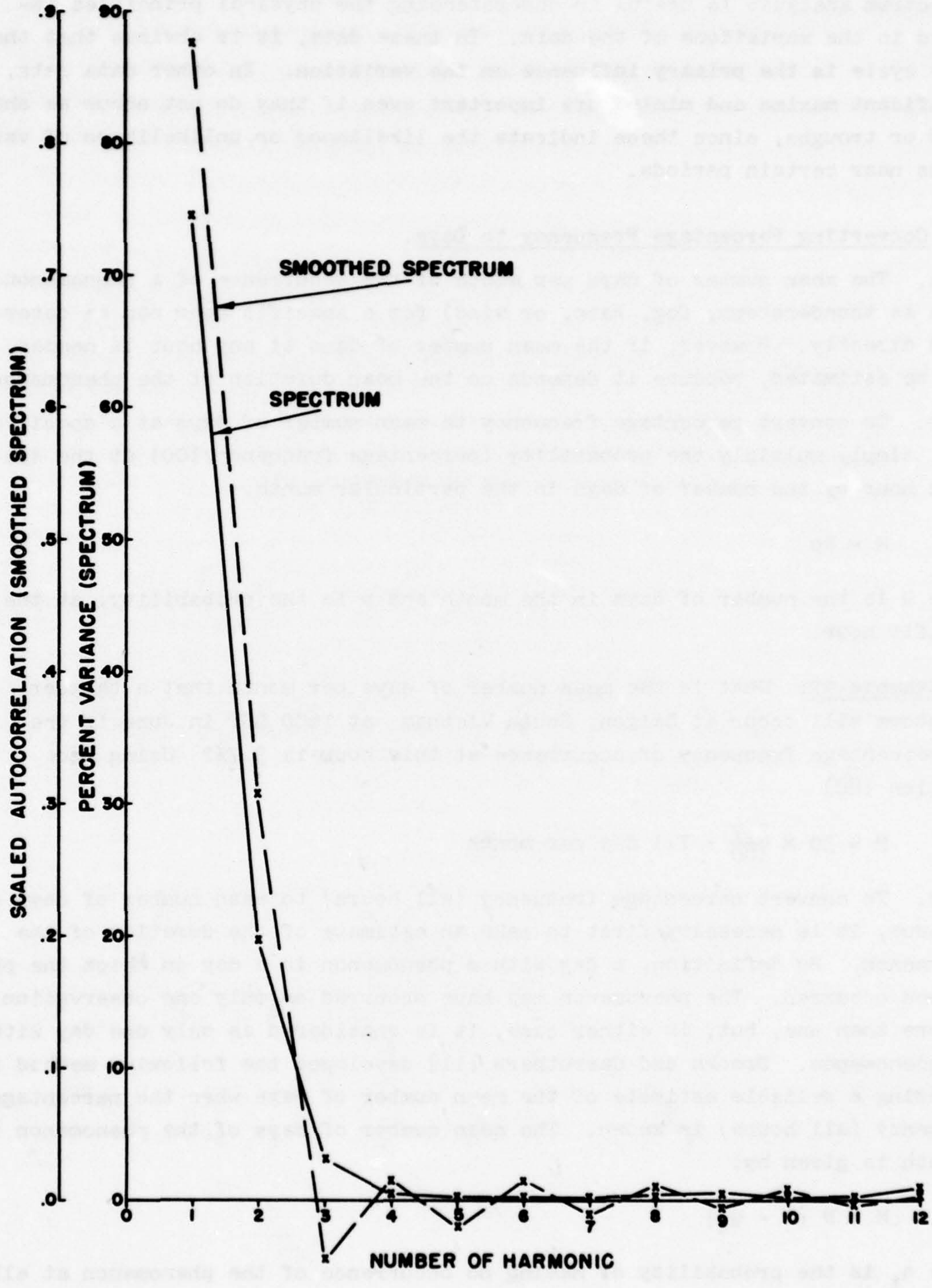


Figure 13. Spectrum of Harmonic Variances and Smoothed Spectrum.

A spectrum analysis is useful in understanding the physical principles involved in the variations of the data. In these data, it is obvious that the daily cycle is the primary influence on the variation. In other data sets, significant maxima and minima are important even if they do not occur as sharp peaks or troughs, since these indicate the likelihood or unlikelihood of variations near certain periods.

11. Converting Percentage Frequency to Days.

a. The mean number of days per month of the occurrence of a phenomenon (such as thunderstorm, fog, rain, or wind) for a specific hour can be determined directly. However, if the mean number of days at any hour is needed, it must be estimated, because it depends on the mean duration of the phenomenon.

b. To convert percentage frequency to mean number of days at a specific hour, simply multiply the probability (percentage frequency/100) at the specific hour by the number of days in the particular month.

(88) $$M = Np$$

where N is the number of days in the month and p is the probability, at the specific hour.

Example 42: What is the mean number of days per month that a thunderstorm will occur at Saigon, South Vietnam at 1600 LST in June if the percentage frequency of occurrence at this hour is 3.7%? Using Equation (88)

$$M = 30 \times \frac{3.7}{100} = 1.1 \text{ day per month}$$

c. To convert percentage frequency (all hours) to mean number of days at any hour, it is necessary first to make an estimate of the duration of the phenomenon. By definition, a day with a phenomenon is a day on which the phenomenon occurred. The phenomenon may have occurred on only one observation or on more than one, but, in either case, it is considered as only one day with the phenomenon. Brooks and Carruthers [11] developed the following method for providing a reliable estimate of the mean number of days when the percentage frequency (all hours) is known. The mean number of days of the phenomenon in a month is given by:

(89) $$M = N\ (1 - q_t)$$

where q_t is the probability of having no occurrence of the phenomenon at all on any one day, and N is the number of days in the particular month. The problem is to express q_t in terms of the all-hours probability p of the

occurrence of the phenomenon. The probability q_t equals the probability of no occurrence at the first hour times the probability of no independent occurrence on the second and all subsequent observations on that day. The probability of nonoccurrence at the first hour is equal to the all-hours probability of nonoccurrence of the phenomenon q. The probability of an independent occurrence of the phenomenon at the second and subsequent observation hours is equal to $p(1 - a)$, where $1 - a$ is the probability that the phenomenon will not occur on any two successive observations and a is given by:

$$(90) \qquad a = 1 - \frac{1}{D}$$

D is the mean duration of the phenomenon in hours. Since $p(1 - a)$ is equal to the probability of an independent occurrence, $1 - p(1 - a) = q + ap$ is the probability of no independent occurrence on any of the 23 observations following the first hour. Therefore, the probability of having no occurrence at all on any one day is given by:

$$(91) \qquad q_t = q(q + ap)^{23}$$

Substitution of Equation (90) into Equation (91) gives:

$$q_t = q\left[q + \left(1 - \frac{1}{D}\right) p\right]^{23} = q\left(q + p - \frac{p}{D}\right)^{23}$$

and since $p + q = 1$

$$(92) \qquad q_t = q\left(1 - \frac{p}{D}\right)^{23}$$

If Equation (92) is substituted into Equation (89), the resulting equation is

$$(93) \qquad M = N\left[1 - q\left(1 - \frac{p}{D}\right)^{23}\right]$$

Equation (93) may be used to find the mean number of days per month, when the percentage frequency (all hours) is known and an estimate of the mean duration can be made. Figure 14 is the graphical solution of Equation (93); the mean number of days can be determined readily from it.

d. The new USAF Revised Uniform Summary of Surface Weather Observations gives the percentage frequency (all hours) of various weather phenomena and the percentage of days with the same phenomena. These data were used with Equation (93) to derive the mean duration of thunderstorms, rain and/or drizzle, snow and/or sleet, and fog at four USAF bases. Observed mean durations of these phenomena for four airbases were calculated from actual weather observations and compared with the derived values. Table 23 shows agreement to be quite good. Percentage frequencies (all hours) and percentages of days

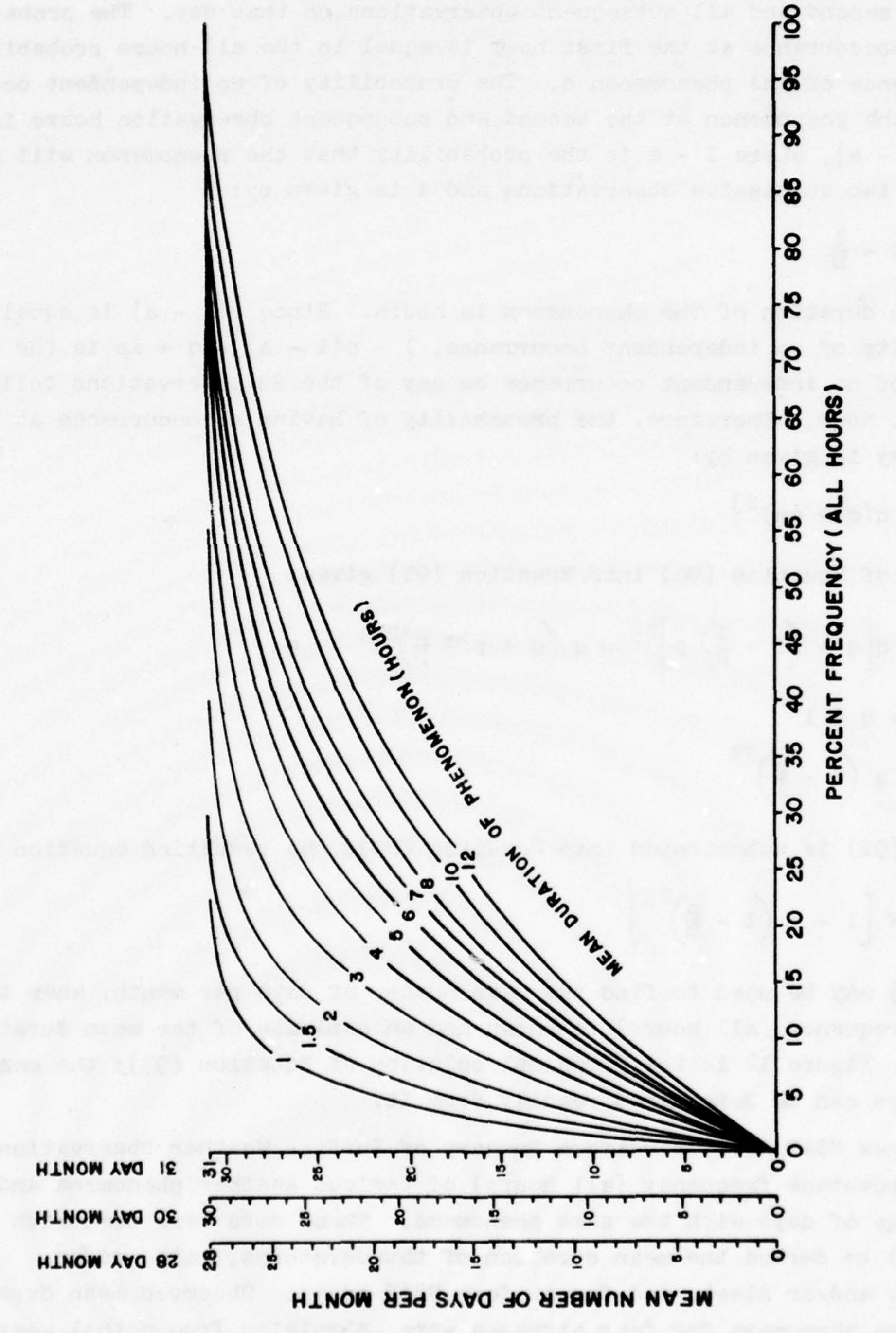


Figure 14. Conversion of Percent Frequency (All Hours) to Mean Number of Days per Month.

TABLE 23

Derived and Observed Mean Duration of Weather Phenomena.

	Thunderstorms		Snow and/or Sleet	
	January and July		January	
Station	Derived	Observed	Derived	Observed
Bergstrom	1.7	1.7	5.6	2.6
Carswell	1.5	2.0	3.8	3.1
Lockbourne	1.8	1.6	5.9	4.0
McGuire	1.6	1.7	3.9	4.2

	Rain and/or Drizzle			
	January		July	
Bergstrom	6.3	4.2	1.9	1.9
Carswell	4.4	3.4	2.5	1.9
Lockbourne	6.7	4.2	2.2	2.0
McGuire	6.2	4.4	3.0	2.6

	Fog			
	January		July	
Bergstrom	10.5	8.1	3.1	1.9
Carswell	9.8	8.9	5.2	3.0
Lockbourne	9.2	9.0	4.0	4.1
McGuire	9.8	8.3	5.2	6.1

of these phenomena were taken from the Revised Uniform Summary of Surface Weather Observations for 63 USAF bases and used with Equation (93) to derive mean durations of weather phenomena at these airbases. Next, averages and standard deviations were calculated; they are given in Table 24. These average values should give good estimates of D, the mean duration, for thunderstorms, rain and/or drizzle, snow and/or sleet, and fog.

Example 43: How many days with thunderstorms can be expected at Miami, Florida during the month of August, if the percentage frequency (all hours) is 4%? From Table 24, the mean duration, D, of thunderstorms is estimated to be 1.6 hours. Consequently, for a percentage frequency of 4% and D of 1.6 hours, Figure 14 gives 15 days with thunderstorms, which is very close to the observed value of 16 days at Miami.

TABLE 24

Mean Duration of Phenomena in Hours at USAF Bases.

Thunderstorms				
Middle Latitude and Tropics	Jan	Apr	Jul	Oct
Average Mean Duration	1.7	1.6	1.6	1.5
Standard Deviation	1.2	0.6	0.4	0.5
No. of Bases Reporting	22	45	49	50

Rain and/or Drizzle				
Arctic and Middle Latitude	Jan	Apr	Jul	Oct
Average Mean Duration	5.0	4.4	2.7	4.2
Standard Deviation	1.3	2.3	1.2	1.1
No. of Bases Reporting	55	55	54	55

Rain and/or Drizzle				
Tropics	Jan	Apr	Jul	Oct
Average Mean Duration	1.5	1.3	1.3	1.3
Standard Deviation	0.6	0.5	0.3	0.3
No. of Bases Reporting	7	7	7	7

Snow and/or Sleet				
Arctic and Middle Latitude	Jan	Apr	Jul	Oct
Average Mean Duration	4.8	3.9	--	3.7
Standard Deviation	1.4	1.9	--	1.9
No. of Bases Reporting	46	29	--	18

Fog				
Arctic, Middle Latitude and Tropics	Jan	Apr	Jul	Oct
Average Mean Duration	8.3	5.3	4.1	5.4
Standard Deviation	2.8	2.0	1.5	2.1
No. of Bases Reporting	59	56	53	57

12. Extreme Value Analysis.

a. Extreme values for wind speed, temperature, precipitation, and snow load constitute a field for the applicati n of extreme value statistics. Many

questions concerning engineering design require knowledge of the extreme value to be expected in a stated number of years. Buildings and antennae must be able to withstand the strongest wind, roofs must withstand the greatest snow load, and dams must withstand the maximum flood anticipated in the lifetime of the structure. In such problems, specified calculated risks may be taken when the likelihood of occurrence of these extremes can be estimated. The statistical theory of extreme values is the method of estimating the extreme values to be expected in a given period, the lifetime of the structure. Different investigators advocate using slightly different theoretical distributions to compute extremes from observed data. The most frequently used theoretical distribution for extreme values is the double-exponential distribution. Gringorten [26] covers the task of selecting an ideal distribution for extreme values. Court [17] demonstrates that the distribution of 5-minute annual extreme winds is approximated exceptionally well by the double-exponential distribution. Gumbel [27] presents an excellent and exhaustive discussion of extreme value theory. Thom [46] discusses applications of the theory to meteorological extremes.

b. The theory applies to the largest (or smallest) values in each of N independent sets of n independent observations drawn from the same population. This parent population must be distributed according to some exponential law so that it is unlimited but tends to zero as the variable increases or decreases. The distribution must also possess all moments. The fundamental theorem of the theory of extreme value is: in a set of N independent extremes $x_1, x_2, x_3, \ldots, x_N$, each being the extreme of n observations of an unlimited, exponentially-distributed variable, as both N and n grow large, the cumulative probability that any of these N extremes will be less than any chosen quantity, x, approaches the double-exponential expression

$$(94) \quad F(x) = \exp\left[-e^{-a(x-\tilde{x})}\right]$$

where: x = some value of the variable, and $\tilde{x}$ = most frequent value (mode) of the set of extremes.

c. The two values a and $\tilde{x}$ are estimated by the theory of least squares from the data of the sample, using two theoretical quantities:

$$(95) \quad a = \frac{\sigma_N}{s_x}$$

and

$$(96) \quad \tilde{x} = \bar{x} - s_x\left(\frac{\bar{Y}_N}{\sigma_N}\right)$$

The mean is $\overline{x}$ and the standard deviation of the set of extremes (sample) is s_x, while the mean $\overline{Y}_N$ and standard deviation σ_N of the theoretical variate depend only on the sample size N. Since the double-exponential form of the basic Equation (94) imposes difficulties in computation and analysis, it is reduced to linear form by taking the double logarithm of both sides. The new variate, $y(x) = -\ln[-\ln F(x)]$, is called the reduced variate:

$$y = a(x-\tilde{x})$$

Solved for x, this equation becomes:

$$x = \tilde{x} + \frac{y}{a}$$

After substitution of Equations (95) and (96), this expression becomes

$$(97) \qquad x = \overline{x} + \left(\frac{s_x}{\sigma_N}\right)(y - \overline{Y}_N)$$

This equation gives the expected extreme for any set of N extremes for specified probability of nonoccurrence given by y. Equation (97) is the basic equation for computing various expected extremes and gives the "frequency factor" for the theory of extreme values:

$$(98) \qquad K = \frac{(y - \overline{Y}_N)}{\sigma_N}$$

Since the reduced variate y is the double logarithm of the probability and $\overline{Y}_N$ and σ_N depend only on the sample size, K can be tabulated for use. Table 25 presents values of K for various probabilities, F(x), and various sample sizes, N. If additional values are required for a specific problem, they can be easily computed from Equation (98) and the reduced variate equation. The "general formula" for the "line of expected extremes" is

$$(99) \qquad x = \overline{x} + K(s_x)$$

where x is the expected extreme, whose probability of not being equaled is $F(x) = \exp(-e^{-y})$, and $\overline{x}$ and s_x are the mean and standard deviation from the available sample.

d. The term "return period" is used frequently in extreme value analysis. By definition:

> An event that happens A times in N trials has a relative frequency of occurrence of A/N and a return period of RP = N/A.

The return period, or reciprocal of the relative frequency, is therefore the

TABLE 25

Values of K for Various Probabilities and Sample Sizes.

Probability: F(x)			.999	.99	.975	.95	.80
Return Period:			1000	100	40	20	5
$y = -\ln[-\ln F(x)]$:			6.90726	4.60015	3.67625	2.97020	1.49994
N	$\bar{y}_N$	σ_N	Values of K				
15	.5128	1.0206	6.265	4.005	3.100	2.408	0.967
16	.5157	1.0316	6.196	3.959	3.064	2.379	0.954
17	.5181	1.0411	6.137	3.921	3.033	2.355	0.943
18	.5202	1.0493	6.087	3.888	3.008	2.335	0.934
19	.5220	1.0566	6.043	3.860	2.985	2.317	0.926
20	.5236	1.0628	6.006	3.836	2.966	2.302	0.919
21	.5252	1.0696	5.967	3.810	2.946	2.286	0.911
22	.5268	1.0754	5.933	3.788	2.929	2.272	9.905
23	.5283	1.0811	5.900	3.766	2.912	2.259	0.899
24	.5296	1.0864	5.870	3.747	2.896	2.247	0.893
25	.5309	1.0914	5.842	3.728	2.882	2.235	0.888
26	.5816	1.0961	5.816	3.711	2.869	2.224	0.883
27	.5332	1.1004	5.792	3.696	2.856	2.215	0.879
28	.5343	1.1047	5.769	3.681	2.844	2.205	0.874
29	.5353	1.1086	5.748	3.667	2.833	2.196	0.870
30	.5362	1.1124	5.727	3.653	2.823	2.188	0.866
40	.5436	1.1413	5.576	3.554	2.745	2.126	0.838
50	.5485	1.1607	5.478	3.491	2.695	2.086	0.820
60	.5521	1.1747	5.410	3.446	2.660	2.058	0.807
70	.5548	1.1854	5.359	3.413	2.633	2.038	0.797
80	.5569	1.1938	5.319	3.387	2.613	2.022	0.790
90	.5586	1.2007	5.287	3.366	2.597	2.009	0.784
100	.5600	1.2065	5.261	3.349	2.583	1.998	0.779
1000	.5745	1.2685	4.992	3.174	2.445	1.889	0.730
∞	.5772	1.2826	4.936	3.137	2.416	1.866	0.719

<u>AVERAGE</u> interval between recurrences of the event in a particular series of trials. Expressed in terms of probability, the return period is

$$RP = \frac{1}{1 - F(x)} \quad (100)$$

where F(x) has been previously defined (page 4-53) and 1 - F(x) is the relative frequency, f(x), of the event. The return period is often misunderstood and should not be used except as a tool in computing calculated risks. <u>Calculated risk</u> is defined as the probability of at least one occurrence of an event during a specified time interval. Given the variable x, a calculated risk of some value of x can be stated for one year, five years, ten years, or

longer. Consider the binomial theorem (page 3-16, Chapter 3). If n is any positive integer and p and q are any numbers, then

$$(p + q)^n = \sum_{r=0}^{n} \binom{n}{r} p^{n-r} q^r$$

If p is the probability of some event and q, its complement, and the only possible events are p and q, then the probability that p will occur in all n trials is

$$\sum_{r=0}^{n} \binom{n}{0} p^{n-0} q^0 = p^n$$

It was seen earlier that F(x) is the probability that the event x will not occur. The probability that x will not occur in n trials is $[F(x)]^n$. Hence, the probability of at least one occurrence (calculated risk) of the event x is.

$$(101) \qquad f(x)_n = 1 - [F(x)]^n$$

To express this in terms of the return period, it is only necessary to rearrange Equation (100) to give

$$(102) \qquad F(x) = \frac{(RP - 1)}{RP}$$

Substitution into Equation (101) yields

$$(103) \qquad f(x)_n = 1 - \left[\frac{(RP - 1)}{RP}\right]^n$$

If $f(x)_n$ is fixed in advance, then RP may be found by conversion of Equation (103) to

$$(104) \qquad RP = \frac{1}{\{1 - [1 - f(x)_n]^{1/n}\}}$$

Solution of Equation (104) gives necessary return periods for various values of n and $f(x)_n$. Table 26 gives these values for several combinations of $f(x)_n$ and n.

e. Figure 15 illustrates the special "extreme probability paper" prepared by Gumbel [27] and modified by Court [17], in which the cumulative probability, F(x), is plotted on the left ordinate and the variate, x, on the abscissa. The right ordinate is a quasi logarithmic scale for the return period. On this paper a double-exponential distribution appears as a straight line. Extreme probability paper is identical in function and use to other probability papers; observations are plotted on it by rank and magnitude.

TABLE 26

Return Periods for Specified Calculated Risks and Design Life.

Calculated Risk $[f(x)_n]$	Design Life in Years (n)*											
	1	2	3	4	5	10	15	20	25	30	35	40
.01	100	199	299	398	498	996	1493	1991	2488	2986	3483	3981
.02	50	99	149	198	248	495	743	990	1238	1485	1733	1980
.03	33	66	99	132	165	329	493	657	821	985	1150	1314
.04	25	49	74	98	123	245	368	490	613	735	858	980
.05	20	39	59	78	98	195	293	390	488	585	683	780
.10	10	19	29	38	48	95	143	190	238	285	333	380
.15	7	13	19	25	31	62	93	124	154	185	216	247
.20	5	9	14	18	23	45	68	90	113	135	157	180
.25	4	7	11	14	18	35	53	70	87	105	122	140
.30	3	6	9	12	15	29	43	57	71	85	99	113
.35	3	5	7	10	12	24	35	47	59	70	82	93
.40	3	4	6	8	10	20	30	40	49	59	69	79
.45	2	4	6	7	9	17	26	34	42	51	59	67
.50	2	3	5	6	8	15	22	29	37	44	51	58

* Return Period (RP) = $\dfrac{1}{\{1 - [1 - f(x)_n]^{1/n}\}}$

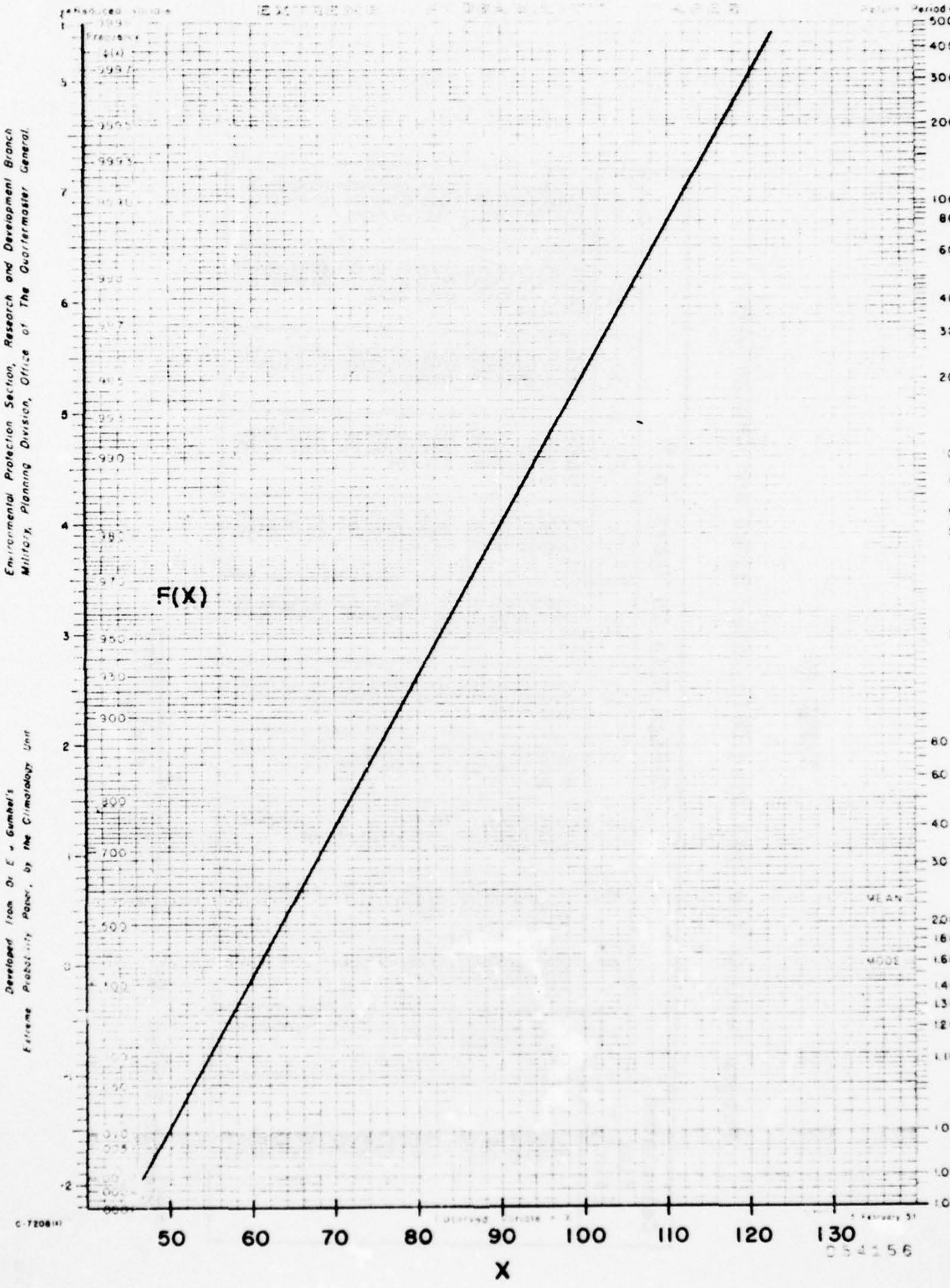


Figure 15. Example of Double-Exponential Distribution on Extreme Probability Paper.

Each extreme is plotted at an abscissa corresponding to its value and at an ordinate, on the double logarithmic scale, corresponding to its cumulative rank divided by N + 1. The abscissa for the 12th largest extreme of 24 extreme values is 13/25 or 0.52. These points should approximate a straight line if the set of extremes follows the theory of extreme values.

Example 44: A structure is to be built at Argentia NAS, Newfoundland. The structure is planned for a useful life of 10 years. A calculated risk of 10% that the design wind speed will be equaled, or exceeded, during the 10 years (design life) is specified. What is the wind speed that meets these criteria?

Solution: From climatic data shown in Table 27, the following annual maximum winds are available for Argentia NAS (values are for 30 feet above the ground):

TABLE 27

Peak Gust Wind Data for Argentia NAS.

Year	Peak Gust (kts)	Year	Peak Gust (kts)	Year	Peak Gust (kts)
1941	73	1949	72	1957	78
1942	76	1950	77	1958	74
1943	73	1951	79	1959	90
1944	72	1952	80	1960	87
1945	71	1953	74	1961	69
1946	77	1954	77	1962	67
1947	62	1955	91	1963	72
1948	75	1956	83	--	--

Compute the mean, $\bar{x}$, and the standard deviation, s_x, of the above peak gust observations (N = 23):

$$\bar{x} = \frac{\Sigma X}{N} = \frac{1749}{23} = 76.04 \text{ knots}$$

$$s_x{}^2 = \frac{\Sigma X^2 - \bar{X}(\Sigma X)}{N - 1} = \frac{134{,}049 - 132{,}994}{22} = 47.95$$

or

$$s_x = 6.92 \text{ knots}$$

The estimate of the extreme value, x, for a given probability of being equaled or exceeded during any one year is given by $x = \overline{x} + s_x K$, where $\overline{x}$ and s_x are the mean and standard deviation of the available sample. K is the frequency factor; it varies with the probability level and sample size. From Table 25, select at least three probability levels to compute the "line of expected extremes."

Probability Level	K (N = 23)
.999	5.900
.99	3.766
.95	2.259
.80	0.899

From the above equation

$$x_{.999} = \overline{x} + s_x\ (5.900) = 76.04 + 6.92\ (5.900) = 116.9 \text{ knots}$$

$$x_{.99} = 76.04 + 6.92\ (3.766) = 102.1 \text{ knots}$$

$$x_{.95} = 76.04 + 6.92\ (2.259) = 91.7 \text{ knots}$$

$$x_{.80} = 76.04 + 6.92\ (0.899) = 82.3 \text{ knots}$$

Figure 15 is the plot of the "line of expected extremes." The straight line on the extreme probability paper in Figure 16 gives the calculated distribution of annual extremes of the peak gust wind speed. From this curve, read the calculated risk for any wind speed during any one year. Determine the wind speed for a calculated risk of 10% during 10 years. The right ordinate of the extreme probability paper is expressed in return period. Notice that the 1% risk (.99 probability of nonoccurence) value is equivalent to a 100-year return period. From Table 26, which gives return periods for various calculated risks during specified number of years, select the necessary return period for 10% risk during 10 years. This value is 95 years. The wind speed value corresponding to a 95-year return period is 102 knots. This value (102 knots) has a calculated risk of 10% of being equaled, or exceeded, at least once during any 10-year period. This is the required design wind. The assumption was made that the annual peak wind gusts are double-exponentially distributed. To determine how effectively the annual peak gusts fit the assumed distribution, plot the observed values using the Gumbel plotting rule, that is, the observations are arranged in ascending order and each cumulative rank (lowest to highest) is divided by N + 1 to give the cumulative probability of nonoccurrence. These points are also shown in Figure 16. The plot of these data on extreme probability

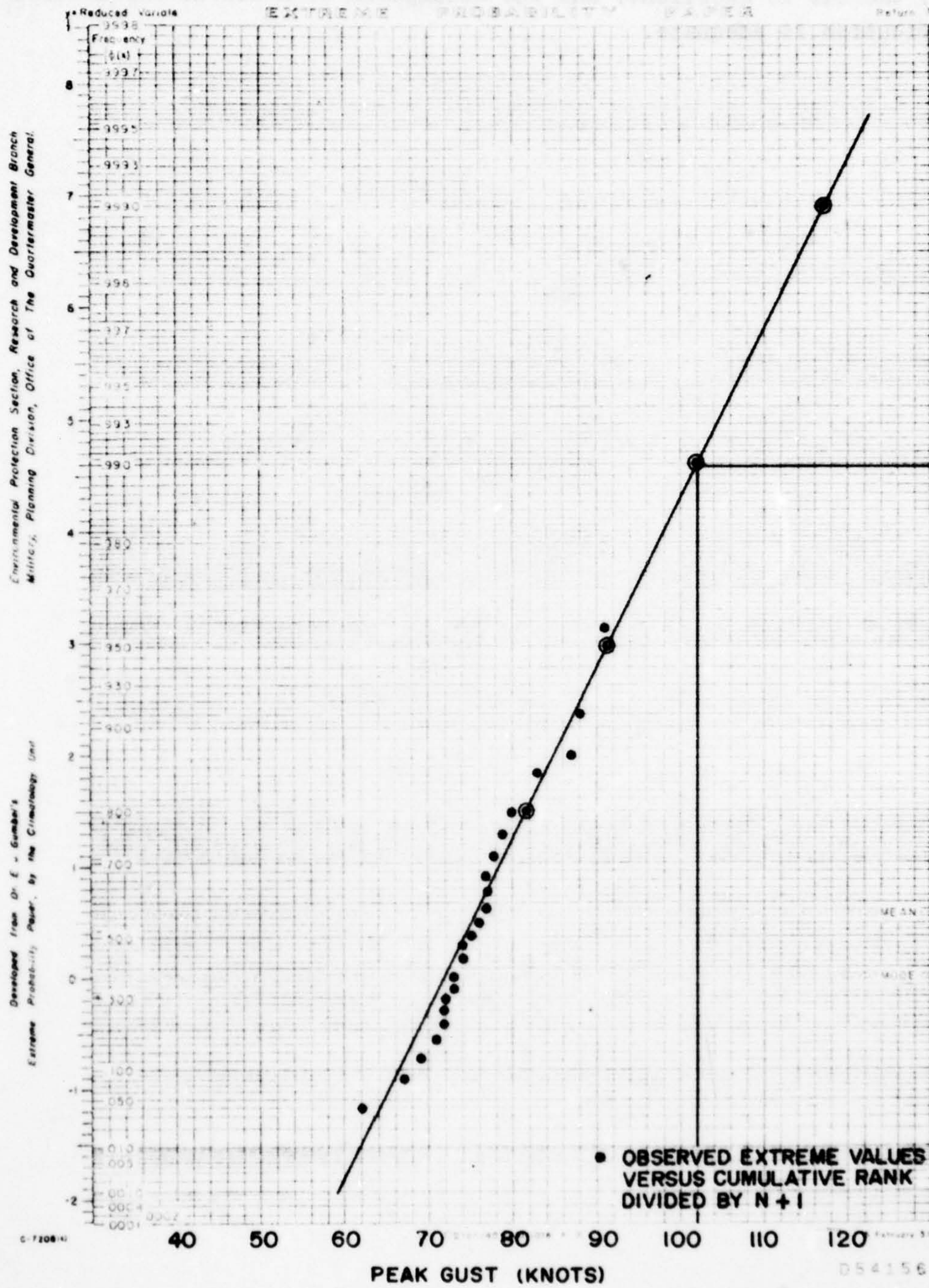


Figure 16. Argentia NAS, Newfoundland, Distribution of Annual Peak Gust Wind Speed.

paper (Figure 16) indicates that these data approximate a straight line, the fit is excellent, and the assumption of a double-exponential distribution is adequate.

Chapter 5

WIND DISTRIBUTIONS

1. Introduction.

One of the main problems of the applied climatologist has been, and continues to be, the determination of wind distribution (speed and direction) through a given layer or at a given pressure surface. To provide adequate service in this area, the climatologist must be familiar with certain standard problem-solving techniques. This chapter deals with the statistical model of the wind distribution at a given level or pressure surface; this model is known generally as the elliptical normal distribution. A special case of this distribution is referred to as the circular normal distribution. One facet of the wind distribution problem is the ballistic wind, defined as a fictitious, uniform wind extending from the ground to bombing altitude, and which is determined in such a way that its effect on the bomb is the same as the variable winds actually encountered. The vector correlations between winds at two levels are the links between the distribution of the ballistic wind and the distribution of wind at a number of levels in the upper air.

2. Circular Normal Wind Distribution.

a. The frequency distribution known as the circular normal distribution has been used frequently to represent the climatological distribution of wind vectors in the upper air over a given location or area. It requires that wind observations form a homogeneous set or, in other words, are drawn from the same population [11]. Figure 17 illustrates the variables that are involved in the development of the circular normal distribution. V_R is the mean resultant vector of the sample, V is any wind observation in the sample, v is the vector difference between V_R and V, θ is the angle between v and V_R, and x and y are the components of v along and at right angles to V_R, respectively.

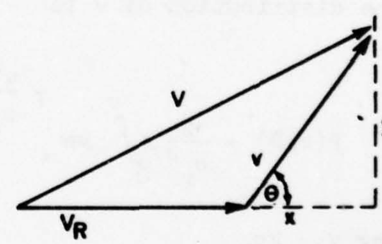


Figure 17. Variables in Circular Normal Distribution.

b. The circular normal distribution requires that the following three conditions exist: (1) x and y are normally distributed, (2) the variances of x and y are equal, and (3) x and y are statistically independent. These conditions determine the form of the joint frequency distribution of x and y. It is the product of two normal distributions.

$$\text{(105)} \quad P(x,y) = \frac{1}{2\pi \sigma_x \sigma_y} \int_{-x}^{x} \int_{-y}^{y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^{\ 2}} + \frac{y^2}{\sigma_y^{\ 2}}\right)} dx\, dy$$

This equation can be transformed into a polar coordinate form with v and θ as the variables. From the stipulated conditions and Figure 17, the following relations hold

$$v^2 = x^2 + y^2$$

$$x = v \operatorname{Cos} \theta, \quad y = v \operatorname{Sin} \theta$$

and

$$dx\, dy = J\left(\frac{x, y}{v, \theta}\right) dv\, d\theta = \begin{vmatrix} \operatorname{Cos} \theta & \operatorname{Sin} \theta \\ -v \operatorname{Sin} \theta & v \operatorname{Cos} \theta \end{vmatrix} dv\, d\theta = v\, dv\, d\theta$$

Therefore, the joint distribution of v and θ is

$$P(v,\theta) = \frac{1}{\pi \sigma_v^{\ 2}} \int_0^v \int_0^\theta v e^{-\frac{v^2}{\sigma_v^{\ 2}}} dv\, d\theta$$

where σ_v is the vector standard deviation. If we integrate this over θ from 2π, the distribution of v is

$$\text{(106)} \quad P(v,\theta) = \frac{2}{\sigma_v^{\ 2}} \int_0^v v e^{-\frac{v^2}{\sigma_v^{\ 2}}} dv = 1 - e^{-\frac{v^2}{\sigma_v^{\ 2}}}$$

Letting $v = k\sigma_v$

$$P(k\sigma_v) = 1 - e^{-k^2}$$

This distribution can be represented as a family of concentric circles centered at the end of the vector V_R and with the radius $k\sigma_v$ (Figure 18). There is a frequency for each circle; it is the frequency that the ends of the wind vectors will be within the circle of radius $k\sigma_v$. The percent probabilities for

various values of the multiplier k, where $k_c = [-\ln(1-P)]^{1/2}$, are:

k =	.32	.47	.60	.71	.83	.96	1.10	1.27	1.52	2.15
P(%) =	10	20	30	40	50	60	70	80	90	99

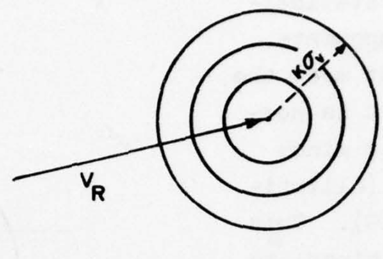


Figure 18. Circular Normal Distribution Probabilities.

3. Test for Circularity.

a. In order for the analyst to know when to accept a distribution as circular, he needs a method to test the significance of the ellipticity of the distribution. Mauchly [36] defines an ellipticity statistic, L_e, by:

$$(107) \quad L_e = \frac{2\,\sigma_x \sigma_y \sqrt{1 - r_{xy}^{\,2}}}{\left(\sigma_x^{\,2} + \sigma_y^{\,2}\right)}$$

In Equation (107), σ_x and σ_y are the standard deviations of the distribution of the individual zonal and meridional components, respectively, and r_{xy} is the correlation coefficient of these zonal and meridional components. For a perfect circular distribution, $\sigma_x = \sigma_y$, $r_{xy} = 0$, and $L_e = 1$; otherwise, L_e is less than 1.

b. The probability of obtaining a value of L_e as small as the value found by using Equation (107) in a sample of N independent observations drawn from a population in which $L_e = 1$ is shown by Mauchly to be L_e^{N-2}. If the 5% level of significance is adopted, Brooks and Carruthers [11] state that the analyst should accept the distribution as circular, provided that

$$\frac{2}{L_e} < 2 \times 20^{1/(N-2)}$$

4. Elliptical Normal Wind Distribution.

a. The circular normal distribution model has proved extremely useful in a variety of applications as an approximation of the true distribution. However, many of the actual distributions of the upper winds do not satisfy the criteria of equal variance for the components of v or that of statistical independence of the components. If the distribution does not meet the criteria for circularity, it is possible to represent the upper winds with the basic noncircular (elliptical) distribution (Figure 19). This is the general form of the bivariate normal distribution; it requires three parameters (the variances of the two components and the correlation coefficient) in place of the standard vector deviation or its square (vector variance). This joint distribution of the x- and y-components is:

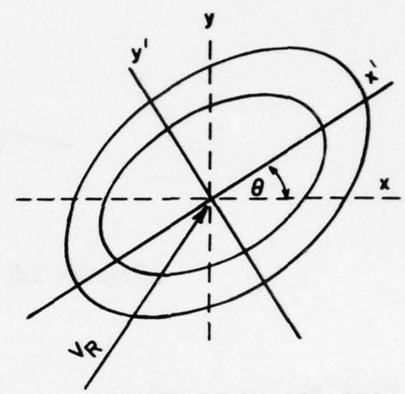


Figure 19. Elliptical Normal Distribution Probabilities.

$$(108) \quad P(x,y) = \frac{1}{2\pi\, \sigma_x \sigma_y (1-r^2)^{1/2}} \int_{-x}^{x} \int_{-y}^{y} e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_x{}^2} - \frac{2rxy}{\sigma_x \sigma_y} + \frac{y^2}{\sigma_y{}^2}\right)} dx\, dy$$

For this model of the upper wind, the projection of the frequency surface onto the xy-plane will be represented by a family of ellipses.

b. The application of this elliptical normal distribution to problems involving wind vectors is very difficult when the distribution is expressed in the form of Equation (108). However, difficulties can be avoided by changing the form of the equation to one that permits the analyst to compute readily the probability of a wind vector being in a given ellipse.

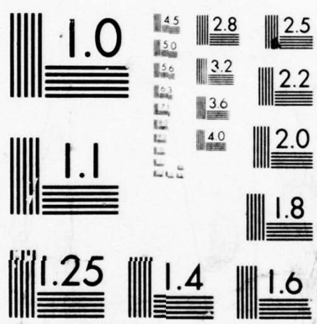
c. To develop the equation for the probability ellipse, let x and y be the components of the departure from the mean resultant vector (V_R) in the conventional axes, east and north, respectively, and let x' and y' be the corresponding departures referred to the axes rotated counterclockwise through θ. The origin is at the extremity of the mean resultant wind, $x' = x \text{ Cos } \theta + y \text{ Sin } \theta$, and $y' = y \text{ Cos } \theta - x \text{ Sin } \theta$.

(1) From the preceding equations, Scott [43] deduces that

AD-A052 579 AIR WEATHER SERVICE SCOTT AFB ILL F/G 4/2
GUIDE FOR APPLIED CLIMATOLOGY.(U)
NOV 77
UNCLASSIFIED AWS-TR-77-267 NL

2 OF 2
AD
A052579

END
DATE
FILMED
5-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

$$
(109)\quad \begin{cases} \sigma_{x'}^2 = \sigma_x^2 \operatorname{Cos}^2\theta + \sigma_y^2 \operatorname{Sin}^2\theta + 2r\sigma_x\sigma_y \operatorname{Sin}\theta \operatorname{Cos}\theta \\ \sigma_{y'}^2 = \sigma_y^2 \operatorname{Cos}^2\theta + \sigma_x^2 \operatorname{Sin}^2\theta - 2r\sigma_x\sigma_y \operatorname{Sin}\theta \operatorname{Cos}\theta \\ r'\sigma_{x'}\sigma_{y'} = r\sigma_x\sigma_y (\operatorname{Cos}^2\theta - \operatorname{Sin}^2\theta) + (\sigma_y^2 - \sigma_x^2) \operatorname{Sin}\theta \operatorname{Cos}\theta \end{cases}
$$

where r and r' are the correlation coefficients of the non-primed and primed components, respectively. The σ_x, σ_y, and r are reported in standard upper-wind summaries. With these values known, it is possible to select an angle so that r' = 0 and solve for $\sigma_{x'}$ and $\sigma_{y'}$. Letting r' equal to zero

$$0 = 2r\sigma_x\sigma_y \operatorname{Cos}\theta + (\sigma_x^2 - \sigma_y^2) \operatorname{Sin} 2\theta$$

or

$$\operatorname{Tan} 2\theta = \frac{2r\sigma_x\sigma_y}{(\sigma_x^2 - \sigma_y^2)}$$

and

$$(110)\quad \theta = \tfrac{1}{2} \operatorname{ARCTAN} \left[\frac{2r\sigma_x\sigma_y}{(\sigma_x^2 - \sigma_y^2)}\right]$$

Equation (110) gives the rotation angle θ for substitution into Equation (109) to determine $\sigma_{x'}$ and $\sigma_{y'}$, standard deviations of wind components along the major and minor axes of the distribution, respectively. These two standard deviations ($\sigma_{x'}$ and $\sigma_{y'}$) are given as σ_a and σ_b in the standard upper-wind summaries and henceforth in this chapter, $\sigma_{x'}$ becomes σ_a and $\sigma_{y'}$ becomes σ_b.

(2) The following method provides an easier means for computing σ_a and σ_b:

$$\begin{bmatrix} \sigma_x^2 - K & r\sigma_x\sigma_y \\ r\sigma_x\sigma_y & \sigma_y^2 - K \end{bmatrix} = 0$$

The expansion of this determinant is a quadratic equation:

$$(111)\quad K = \frac{\left(\sigma_x^2 + \sigma_y^2\right) \pm \sqrt{\left(\sigma_x^2 + \sigma_y^2\right)^2 - 4\sigma_x^2\sigma_y^2 (1-r^2)}}{2}$$

The solutions are the values of σ_a^2 and σ_b^2. The larger value of K using Equation (111) is the variance (σ_a^2) of the wind components along the major

axis of the distribution and the smaller value is the variance (σ_b^2) of the wind components along the minor axis. Therefore, the positive square roots of σ_a^2 and σ_b^2 are the desired standard deviations, σ_a and σ_b. With $r' = 0$, the x' and y' components can be treated as independent, and the wind distribution can be expressed as the joint distribution of x' and y'.

$$P(x',y') = \frac{1}{2\pi\,\sigma_a\sigma_b} \int_{-a}^{a}\int_{-b}^{b} e^{-\frac{1}{2}\left(\frac{x'^2}{\sigma_a^2} + \frac{y'^2}{\sigma_b^2}\right)}\, dx'\, xy'$$

where $x'^2/a^2 + y'^2/b^2 = 1$ and $a/b = \sigma_a/\sigma_b$. Let $u = a/\sigma_a$ and $v = b/\sigma_b$, then

$$P(u,v) = \frac{1}{2\pi} \int_{-a/\sigma_a}^{a/\sigma_a}\int_{-b/\sigma_b}^{b/\sigma_b} e^{-\frac{1}{2}(u^2+v^2)}\, dx'\, dy'$$

Now transforming to polar coordinates, let $z^2 = u^2 + v^2$, $u = z \text{ Cos } \theta$, and $v = z \text{ Sin } \theta$. Then

$$P(z,\theta) = \frac{1}{2\pi} \int_{0}^{a/\sigma_a}\int_{0}^{\theta} z\, e^{-z^2/2}\, dz\, d\theta$$

Integration of the above equation, when

$$P\left(\frac{x'^2}{a^2} + \frac{y'^2}{b^2} = 1\right)$$

is the probability that the wind vector ends within the ellipse defined by the parenthetical equation, gives:

$$(112)\qquad P\left(\frac{x'^2}{a^2} + \frac{y'^2}{b^2} = 1\right) = 1 - e^{-\frac{a^2}{2\sigma_a^2}}$$

Therefore, the major axis of the ellipse, a, is expressed in multiples of σ_a ($a = k_e\, \sigma_a$). Percent probabilities for various values of the multiplier, k, where $k_e = \{2\, \ell n\, [1/(1 - P)]\}^{\frac{1}{2}}$ are

k_e =	.46	.67	.84	1.01	1.18	1.35	1.55	1.79	2.15	3.03
P(%) =	10	20	30	40	50	60	70	80	90	99

So $a = k_e\, \sigma_a$ and $b = k_e\, \sigma_b$ are used as the major and minor axes of ellipses

that contain the specified percent of the distribution. For example, 60% of the wind vectors end within the ellipse with major axis a = 1.35 σ_a and with minor axis b = 1.35 σ_b.

5. Correlation.

a. The models of the wind distribution at a level or on a pressure surface can be extended to encompass the distribution of ballistic wind. This is feasible since the ballistic wind is a linear combination of winds at a number of levels. However, it is necessary to stop and discuss the correlation of vectors before making this step.

b. Unfortunately, approaches to the correlation of vectors are not as well established as the subject of correlation of scalar quantities. A number of techniques of vector correlations have been proposed; their answers differ and there is not complete agreement as to which is the best approach. Two approaches are developed here, one proposed by Court [17] and the other by Durst [23].

c. Concepts and equations of linear correlation are used in the development of vector correlation. For convenience of reference, some of the necessary equations will be listed without derivation. For simplicity, but with no loss of generalization, variables are treated as departures from the mean and the relation between v and x is expressed as

$$v + q = m(v/x) = ax$$

where v is the dependent variable, x the independent variable, m(v/x) is the empirically derived function for estimating v given x, q is the error of estimate, and a is the regression coefficient. The standard error of estimate $(s_{v/x})$ is obtained by

$$\text{(113)} \quad \frac{1}{N} \sum q^2 = \frac{1}{N} \sum (v - ax)^2 = (s_{v/x})^2$$

where N is the number of cases. The correlation coefficient $(r_{v/x})$ for the linear relation between v and x is given by

$$\text{(114)} \quad \frac{\Sigma q^2}{\Sigma v^2} = \frac{(s_{v/x})^2}{s_v^{\ 2}} = 1 - (r_{v/x})^2$$

where s_v is the standard deviation of v. The equation for linear multivariate correlation is analogous to that for the bivariate correlation. The relation between v and the independent variables x and y is expressed as

$$v + q = m(v/xy) = a_1 x + a_2 y$$

The standard error of estimate is

(115) $$\frac{1}{N} \sum q^2 = \frac{1}{N} \sum (v - a_1 x - a_2 y)^2 = (s_{v/xy})^2$$

The multiple correlation coefficient is given by

(116) $$\frac{\Sigma q^2}{\Sigma v^2} = \frac{(s_{v/xy})^2}{s_v^2} = 1 - (r_{v/xy})^2$$

The multiple correlation coefficient can also be expressed in terms of the correlation coefficients for the various possible pairs of variables:

(117) $$(r_{v/xy})^2 = \frac{[(r_{vx})^2 + (r_{vy})^2 - 2\, r_{vx}\, r_{vy}\, r_{xy}]}{[1 - (r_{xy})^2]}$$

d. In his approach to vector correlation, Court [18] begins by assuming that the relation between two vectors can be expressed in a form analogous to the linear scalar correlations. Assume that the vectors are departures from the mean resultant vector. Thus, the relation is assumed to be $\vec{W} = B\vec{Z}$, where $\vec{W}$ is the dependent vector, $\vec{Z}$ the independent, and B the regression coefficient. $\vec{W}$ and $\vec{Z}$ are vectors at two points in time or space (see Figure 20). The vector error of estimate, $\vec{Q}$ is defined by

(118) $$\vec{W} + \vec{Q} = B\vec{Z}$$

(1) The vector Equation (118) can be written out in full, assuming for convenience that each vector has only two components:

$$\begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} q_u \\ q_v \end{bmatrix} = \begin{bmatrix} b_1 & c_1 \\ b_2 & c_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

This matrix equation can be expressed as a set of linear equations

(119) $$\begin{aligned} u + q_u &= b_1 x + c_1 y \\ v + q_v &= b_2 x + c_2 y \end{aligned}$$

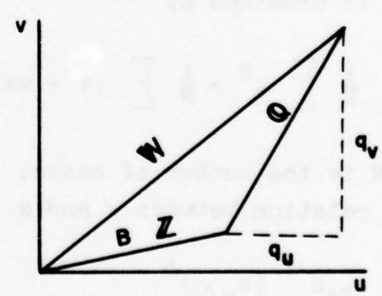


Figure 20. Relationships between Court's Wind Vectors.

Thus, each component of $\vec{W}$ can be estimated by multiple regression on the components of the independent variable $\vec{Z}$.

(2) The vector standard error of estimate $(s_{\vec{W}/\vec{Z}})$ is defined by

$$(120) \qquad (s_{\vec{W}/\vec{Z}})^2 = \frac{1}{N}\sum \vec{Q}^2 = \frac{1}{N}\left(\sum q_u{}^2 + \sum q_v{}^2\right) = (s_{u/xy})^2 + (s_{v/xy})^2$$

(3) The next step is to define the vector correlation coefficient and, from the definition, derive an expression for its value in terms of the correlations between the components of $\vec{W}$ and $\vec{Z}$.

$$(121) \qquad 1 - (r_{\vec{W}/\vec{Z}})^2 = \frac{(s_{\vec{W}/\vec{Z}})^2}{s_{\vec{W}}{}^2} = \frac{(s_{u/xy})^2 + (s_{v/xy})^2}{s_u{}^2 + s_v{}^2}$$

Using Equation (116),

$$(r_{\vec{W}/\vec{Z}})^2 = 1 - \frac{s_u{}^2\,[1 - (r_{u/xy})^2] + s_v{}^2\,[1 - (r_{v/xy})^2]}{s_u{}^2 + s_v{}^2}$$

$$= \frac{s_u{}^2 + s_v{}^2 - s_u{}^2\,[1 - (r_{u/xy})^2] - s_v{}^2\,[1 - (r_{v/xy})^2]}{s_u{}^2 + s_v{}^2}$$

$$(r_{\vec{W}/\vec{Z}})^2 = \frac{s_u{}^2\,(r_{u/xy})^2 + s_v{}^2\,(r_{v/xy})^2}{s_u{}^2 + s_v{}^2}$$

From Equation (117) it is possible to re-write the multiple correlation coefficients to express a vector correlation coefficient in terms of the standard deviations of the components of $\vec{W}$ and the correlation coefficients of pairs on components of the two wind vectors.

$$(122) \qquad (r_{\vec{W}/\vec{Z}})^2 = \frac{s_u{}^2\,[(r_{ux})^2 + (r_{uy})^2 - 2\,r_{ux}\,r_{uy}\,r_{xy}]}{s_{\vec{W}}{}^2\,[1 - (r_{xy})^2]} + \frac{s_v{}^2\,[(r_{vx})^2 + (r_{vy})^2 - 2\,r_{vx}\,r_{vy}\,r_{xy}]}{s_{\vec{W}}{}^2\,[1 - (r_{xy})^2]}$$

Assuming that both $\vec{W}$ and $\vec{Z}$ conform to a circular normal distribution, Equation (122) becomes

$$(123) \qquad (r_{\vec{W}/\vec{Z}})^2 = \frac{[(r_{ux})^2 + (r_{vy})^2]}{2} + \frac{[(r_{uy})^2 + (r_{vx})^2]}{2}$$

The first term on the right side measures the tendency for the two vectors to be parallel and the second term measures the rotation of one vector from the other.

e. In his approach to vector correlation, Durst [23] deals with vectors that are departures of the wind vector from the mean resultant wind and uses $\vec{W}$ and $\vec{Z}$ as noted above. The vector correlation coefficient is composed of two parts, one measuring the parallelism between the two vectors and the other measuring the amount of turn. The analyst can conceive of a relation between $\vec{W}$ and $\vec{Z}$ in which $\vec{W}$ tends to be parallel to $\vec{Z}$ and to be proportional to it in magnitude. The closeness of such a relation can be measured by a form of correlation coefficient that is called the "stretch correlation coefficient" and is expressed as

$$(124) \quad \rho_s = \frac{\Sigma\, \vec{W}\cdot\vec{Z}}{[(\Sigma\vec{W}^2)(\Sigma\vec{Z}^2)]^{1/2}}$$

Other complicated relations can be considered, particularly the situation in which $\vec{W}$ tends to have a direction rotated by an angle α from $\vec{Z}$. The angle α is called the angle of turn and the part of the correlation coefficient that measures the effect of turn is

$$(125) \quad \rho_t = \frac{\Sigma\, |\vec{W}\times\vec{Z}|}{[(\Sigma\vec{W}^2)(\Sigma\vec{Z}^2)]^{1/2}}$$

The total correlation coefficient is composed of both of these (stretch and turn) relationships:

$$(126) \quad \rho = \rho_s + \rho_t$$

The best value of α can be expressed as

$$(127) \quad \alpha = \mathrm{Tan}^{-1} \frac{\Sigma\, |\vec{W}\times\vec{Z}|}{\Sigma(\vec{W}\cdot\vec{Z})}$$

To evaluate the stretch and turn parts of the correlation coefficient, wind vectors can be expressed in terms of their components and dot and cross operations can be performed on these component vectors. Component axes are the same as those used in the previous sections and are shown in Figure 20. Substituting the components into Equation (124),

$$\rho_s = \frac{\Sigma(\vec{u} + \vec{v})\cdot(\vec{x} + \vec{y})}{[\Sigma(\vec{u} + \vec{v})^2\ \Sigma(\vec{x} + \vec{y})^2]^{1/2}} = \frac{\Sigma(ux + vy)}{[\Sigma(u^2 + v^2)\ \Sigma(x^2 + y^2)]^{1/2}}$$

and

(128) $$\rho_s = \frac{s_u \, s_x \, r_{ux} + s_v \, s_y \, r_{vy}}{s_{\vec{W}} \, s_{\vec{Z}}}$$

At this point, assume that the distribution of wind vectors has a circular normal distribution from which it follows that

$$\Sigma\, u^2 = \Sigma\, v^2$$

$$\Sigma\, x^2 = \Sigma\, y^2$$

Using these equalities and separating the term on the right into two terms,

(129) $$\rho_s = \tfrac{1}{2}\left[\frac{\Sigma\, ux}{(\Sigma u^2 \; \Sigma x^2)^{1/2}} + \frac{\Sigma\, vy}{(\Sigma v^2 \; \Sigma y^2)^{1/2}}\right]$$

The two terms within the brackets are the correlation between u and x, and between v and y, respectively. From this, it is evident that the stretch vector correlation coefficient is the average of the correlation between both pairs of components

(130) $$\rho_s = \tfrac{1}{2}\,(r_{ux} + r_{vy})$$

The procedure for evaluating the turn correlation coefficient is similar to that used with the stretch correlation. Substituting the components into Equation (125),

$$\rho_t = \frac{\Sigma(\vec{u} + \vec{v}) \times (\vec{x} + \vec{y})}{[\Sigma(\vec{u} + \vec{v})^2 \; \Sigma(\vec{x} + \vec{y})^2]^{1/2}} = \frac{\Sigma(vx - uy)}{[\Sigma(u^2 + v^2) \; \Sigma(x^2 + y^2)]^{1/2}}$$

(131) $$\rho_t = \frac{s_x \, s_v \, r_{vx} - s_u \, s_y \, r_{uy}}{s_{\vec{W}} \, s_{\vec{Z}}}$$

Again, assuming a circular normal distribution and separating the term on the right into two terms,

(132) $$\rho_t = \tfrac{1}{2}\left[\frac{\Sigma\, vx}{(\Sigma v^2 \; \Sigma x^2)^{1/2}} - \frac{\Sigma\, uy}{(\Sigma u^2 \; \Sigma y^2)^{1/2}}\right]$$

where terms in the brackets are the correlation coefficients. Thus,

(133) $$\rho_t = \tfrac{1}{2}\,(r_{vx} - r_{uy})$$

Substituting Equations (130) and (133) into Equation (126), the vector correlation is

$$\rho = \frac{r_{ux} + r_{uy}}{2} + \frac{r_{vx} - r_{uy}}{2}$$

which, assuming the circular normal distribution, is a simplified equation for ρ. However, the substitution of Equations (128) and (131) into Equation (126) gives the value of ρ without the restriction as to type of distribution:

$$(134) \quad \rho = \frac{s_u\, s_x\, r_{ux} + s_v\, s_y\, r_{vy}}{s_{\vec{W}}\, s_{\vec{Z}}} + \frac{s_v\, s_x\, r_{vx} - s_u\, s_y\, r_{uy}}{s_{\vec{W}}\, s_{\vec{Z}}}$$

The inclusion of the turn portion of the correlation adds very little to the numerical value of the vector correlation in most problems [23]. It usually is omitted and only the stretch vector correlation coefficient used.

f. The two equations for the computation of vector correlation developed in previous sections are considerably different from each other even though both contain a stretch term and a turn term. Also, one is not derivable from the other. For comparison, the two equations are repeated:

Court:

$$(r_{\vec{W}/\vec{Z}})^2 = \frac{s_u^2\,[(r_{ux})^2 + (r_{uy})^2 - 2\, r_{ux}\, r_{uy}\, r_{xy}]}{s_{\vec{W}}^2\,[1 - (r_{xy})^2]}$$

$$+ \frac{s_v^2\,[(r_{vx})^2 + (r_{vy})^2 - 2\, r_{vx}\, r_{vy}\, r_{xy}]}{s_{\vec{W}}^2\,[1 - (r_{xy})^2]}$$ (see page 5-9)

Durst:

$$\rho = \frac{s_u\, s_x\, r_{ux} + s_v\, s_y\, r_{vy}}{s_{\vec{W}}\, s_{\vec{Z}}} + \frac{s_v\, s_x\, r_{vx} - s_u\, s_y\, r_{uy}}{s_{\vec{W}}\, s_{\vec{Z}}}$$

The questions of which is the better measure of correlation and whether other methods should be considered cannot be answered until there is considerably more testing and research. However, the question of how well these two coefficients agree has been investigated. Charles [14] has compared both correlation coefficients temporally and spatially; Scott [44] has compared the two over time intervals. Charles used serially complete winds at 500-, 300-, and 100-mb surfaces for a five-year period at 50 stations; Scott used only one station. For the time-lag vector correlation coefficients, $r_{\vec{W}/\vec{Z}}$ is slightly larger than ρ but there is no appreciable gain in using the computationally cumbersome $r_{\vec{W}/\vec{Z}}$ in place of ρ. When these two correlation coefficients are computed for two sets of spatially-separated winds, the agreement is good when ρ is greater than 0.3; but, when ρ is less than 0.3, there is a major difference between the two coefficients. Of course, part of the trouble is that $r_{\vec{W}/\vec{Z}}$ is defined as a positive number and ρ can be either positive or negative. There are two reasons for using the ρ correlation coefficient and both concern

the ease of computation and manipulation rather than rigorous mathematics. The first reason is that ρ is much easier to compute, and statistical parameters required for its computation are more readily available, especially if one uses the stretch vector correlation coefficient as a close approximation. The second reason is that the stretch vector correlation coefficient acts in a manner similar to the scalar correlation coefficient, since it may be extended to the concept of total and partial vector correlation coefficients [23].

6. Ballistic Wind Distribution.

a. A ballistic wind is defined as a fictitious, single wind that is representative of the layer from the ground to bombing altitude and is determined in such a way that its effect on the bomb is the same as the variable winds actually encountered. Mathematically, it is defined as

$$\vec{V}_b = \int_{\text{sfc}}^{\text{bomb alt.}} w(Z)\,\vec{V}(Z)\,dZ$$

where

$$\int_{\text{sfc}}^{\text{bomb alt.}} w(Z)\,dZ = 1$$

$w(Z)$ is the ballistic weighting as a function of altitude, and $\vec{V}(Z)$ is the wind profile as a function of altitude. Of course, the analyst does not usually know or cannot express the wind as a continuous function of altitude and must adopt a summation process in place of an integral. Therefore, the ballistic wind is defined as:

$$(135) \qquad \vec{V}_b = \sum_{i=1}^{k} w_i \vec{V}_i \quad \text{and} \quad \sum_{i=1}^{k} w_i = 1$$

where k is the number of zones chosen to represent the total bombing zone, w_i is the ballistic weight for zone i, and $\vec{V}_i$ is the wind vector for zone i. The mean resultant ballistic wind is the weighted sum of the mean resultant winds for the k levels.

$$(136) \qquad \bar{\vec{V}}_b = \sum_{i=1}^{k} w_i \bar{\vec{V}}_i$$

The bar indicates a mean. The vector variance can be computed from the variance of each of the components of the ballistic wind. The same notation that was used in discussing the wind distribution at a level is used here, except

that subscripts are used to denote a ballistic wind or component and to indicate levels used in computing a ballistic wind. The departure of a ballistic wind from a mean resultant ballistic wind is

$$\vec{U}_b = \vec{V}_b - \vec{\overline{V}}_b = \vec{X}_b + \vec{Y}_b$$

from this it is possible to derive

$$U_b{}^2 = X_b{}^2 + Y_b{}^2$$

and

$$(137) \quad s_b{}^2 = s_{x_b}^2 + s_{y_b}^2$$

where s_b is the vector standard deviation of the ballistic wind and the two terms on the right are the variances of x and y components of the ballistic wind, respectively.

b. The components of the ballistic wind are defined in the same manner as the ballistic wind and

$$(138) \quad x_b = \sum_{i=1}^{k} w_i \, x_i$$

The variance of x_b is

$$(139) \quad s_{x_b}^2 = \frac{1}{N} \sum^{N} x_b{}^2$$

where $\overset{N}{\Sigma}$ is used to indicate the summation over N ballistic winds in the sample. Substituting Equation (138) into Equation (139), we have

$$s_{x_b}^2 = \frac{1}{N} \sum^{N} \left[\sum_{i=1}^{k} w_i \, x_i \right]^2$$

$$s_{x_b}^2 = \frac{1}{N} \sum^{N} \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \, x_i x_j = \sum_{i=1}^{k} \sum_{j=1}^{k} \sum^{N} \frac{1}{N} (w_i w_j \, x_i x_j)$$

$$(140) \quad s_{x_b}^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \, r_{x_{ij}} \, s_{x_i} \, s_{x_j}$$

where $r_{x_{ij}}$ is the correlation between the x components of the wind at levels i and j. In the same manner, the equation for the variance of the y-component of the ballistic wind can be derived, thus

$$(141) \quad s_{y_b}^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \; r_{y_{ij}} \; s_{y_i} \; s_{y_j}$$

c. Although it is possible to compute the ballistic wind, the mean resultant ballistic wind, its variance and variances of the components, there is the question of what to do about the correlation $r_{x_b y_b}$ between the components of the ballistic wind. The correlation coefficient and the variances of the components are necessary for the determination of whether the distribution is circular or elliptical.

d. Of course, if the variances are equal, the assumption of a circular normal distribution is a natural and probably the most practical assumption. Even though the variances of the components are not equal, the analyst may assume that the circular normal distribution is a good approximation and compute the vector standard deviation from Equation (137). The assumption of circularity applies only to the ballistic wind and not to the distribution of winds at the various levels. However, assuming that the distributions at each level are circular normal, it would follow that the ballistic wind distribution is circular normal, since the ballistic wind is a linear combination of the winds at the various levels. If the wind at level i had a circular normal distribution,

$$(142) \quad s_{x_i}^2 = s_{y_i}^2 = \frac{s_{v_i}^2}{2}$$

where s_{v_i} is the vector standard deviation at level i, and

$$s_b{}^2 = s_{x_b}^2 + s_{y_b}^2$$

$$s_b{}^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \; r_{x_{ij}} \; s_{x_i} \; s_{x_j} + \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \; r_{y_{ij}} \; s_{y_i} \; s_{y_j}$$

From Equation (142), this may be rewritten as

$$s_b{}^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \frac{r_{x_{ij}} + r_{y_{ij}}}{2} s_{v_i} s_{v_j}$$

and, since circularity is assumed, the analyst can use Equation (130) for values of stretch correlation coefficients between levels as an approximation of the vector correlation coefficient. Using Equation (130) values,

$$(143) \quad s_b{}^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \, \rho_{ij} \, s_{v_i} s_{v_j}$$

One thing that makes this equation so useful is that Kochanski [30] has investigated the geographical distribution of the stretch vector correlation coefficient and has developed a procedure for estimating its value for most interlevel combinations (i.e., pairs of levels) over the Northern Hemisphere. Interlevel correlations for components are available for only limited areas in the Northern Hemisphere [14].

e. The assumption that the correlation coefficient, $r_{x_b y_b}$, is zero does not prevent the analyst from using the techniques applicable to an elliptical distribution. This assumption merely implies that the major axis of the ellipse is parallel to one of the coordinate axes. In Figure 21, the major axis is parallel to the x-axis and the variance of the x-component is larger than the variance of the y-component. In this situation and the others that follow, all equations given previously for the elliptical distribution apply.

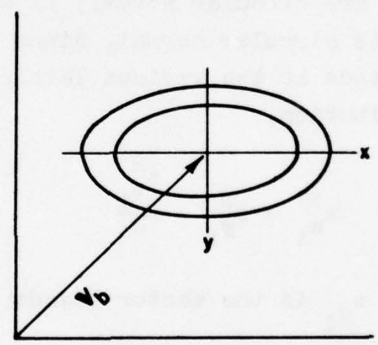


Figure 21. Elliptical Normal Distribution when the Correlation Coefficient Equals Zero.

f. From a study of these data, it might be more appropriate to assume that the major axis of the ellipse lies along the direction of the mean resultant ballistic wind. This assumption implies the assumption of a value for the correlation coefficient for the relation between x and y. Since they are departures from the mean, then

$$Y = bx = r_{x_b y_b} \frac{s_{y_b}}{s_{x_b}} x = (\text{Tan } \psi) \; x$$

where ψ is the slope of the major axis of the ellipse or, in other words, the angle that it makes with the x-axis (see Figure 22). From this, it follows that

$$(144) \quad r_{x_b y_b} = \frac{s_{x_b}}{s_{y_b}} \operatorname{Tan} \psi$$

Of course, this technique is general and may be applied to any other slope of the major axis of the ellipse that one may wish to assume.

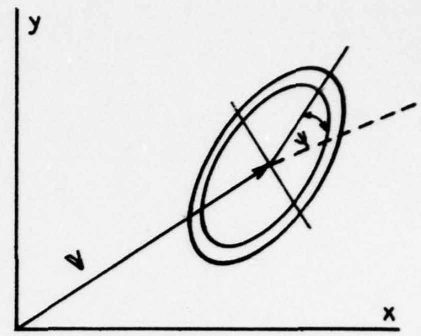


Figure 22. Elliptical Normal Distribution when the Correlation Coefficient is not Equal to Zero.

g. Since the ballistic wind distribution is a combination of the wind distributions at a number of levels, the analyst might assume that orientations of the ellipses at these levels affect the orientation of the ellipses of the ballistic wind distribution. Carrying this a little further, he might assume that the correlation between the components of the ballistic wind is the average of the correlations between the components at the various levels.

$$(145) \quad r_{x_b y_b} = \frac{1}{k} \sum_{i=1}^{k} r_{x_i y_i}$$

where $r_{x_i y_i}$ is the correlation between the components at level i. Then, the analyst merely applies the various equations of the elliptical normal distribution.

h. There are undoubtedly other methods for estimating the correlation between the components of the ballistic wind and determining the better assumption, circularity or ellipticity. This discussion is merely an introduction to the possible methods of handling the distribution of ballistic winds.

Preceding Page Blank - NOT FILMED

Chapter 6

TECHNIQUES IN APPLIED CLIMATOLOGY

1. Introduction.

a. In the solution of applied climatological problems, techniques vary with the degree of complexity. However, three major elements enter into every problem:

(1) Climate, composed of the various weather parameters.

(2) Space, consisting of the surface and upper layers of the atmosphere.

(3) Time, comprising the series and sequence of weather observations.

b. Each of these elements can appear in the problem either as a simple or a complex element. In other words, the climatic factor can involve one weather parameter or many. Space can mean a single point, several points, or an area. Time may enter only in the restricted sense that observations cover an interval of time, or the problem may pose specific time limitations in terms of cumulative effects or conditioning of subsequent events by preceding situations [31].

c. Selected USAF ETAC techniques in applied climatology are included in this chapter as examples of those used to solve the meteorological portions of varied operational, engineering, and design problems of military planners. Many other techniques, methods, and procedures developed at USAF ETAC can be found in the Climatic Methods File (CMF) [9] and various AWS Technical Reports prepared at USAF ETAC.

2. Precipitation, Soil Moisture, and Tractionability Charts.

a. The actual amounts of precipitation and the effects thereof have either direct or indirect application to most military operations, plans, and intelligence. One of the more important aspects concerns knowledge of the moisture content of the soil. This moisture content, directly affecting the trafficability of land areas, is of extreme importance. The amount of soil moisture and the condition of the soil surface cannot be determined adequately by a knowledge of only the amount of water added to the soil by precipitation. Among other factors that must be considered are water loss by evaporation and plant use (evapotranspiration), drainage in and out of the area of concern, and soil types. All of these variable parameters must be considered by the

meteorologist if he is to make a reasonable estimate of soil-moisture content. The following paragraphs describe the technique used at USAF ETAC in dealing with soil-moisture content and tractionability.

b. A method credited to Charles W. Thornthwaite, with modifications and assumptions described herein, is used for soil moisture and tractionability calculations. Precipitation, evapotranspiration, and soil-moisture amounts are given in units of millimeters of water. Changes in soil moisture (ΔSM) during a period depend on the following parameters:

(1) Mean temperature $\overline{T} = \frac{T_{min} + T_{max}}{2}$

(2) Hours of daylight [used to determine potential evapotranspiration (PE)].

(3) Total precipitation (P).

(4) Ratio of the previous soil moisture (at start of period) to the soil moisture at field capacity (previous SM/200).

c. Ten-day periods (decades) are used for soil moisture and tractionability calculations. For USAF ETAC's wide areas of interest, 200 mm of water in the top meter of soil are assumed to provide a full soil bank or to bring soil moisture to field capacity. Therefore, when calculating soil moisture, amounts in excess of 200 mm are considered as surplus, or run-off. In contrast, however, when calculating tractionability, amounts in excess of 200 mm are included to obtain classifications 4 (very moist) and 5 (wet).

d. Daily maximum and minimum temperatures, from which 10-day and monthly mean temperatures are computed, and total daily precipitation amounts are the only meteorological variables that are used in the computation method. It is important to note that these values are reported only in the special phenomena groups (Code 7 groups for Europe and most of Asia, and Code 2 and 3 groups for Southeast Asia).

e. The computational method used at USAF ETAC is basically a bookkeeping procedure. The initial or starting condition of soil moisture for each selected grid point is determined, then the change in soil moisture due to subsequent precipitation, evaporation, and plant use (evapotranspiration) is added to the previous soil-moisture value to obtain the new grid-point value of soil moisture at the end of the period.

(1) All totals of precipitation, means of temperature, and soil-moisture computations are made for fixed grid points. Values of these parameters must be available for all computation points each day for use in the decadal (10-day) computation. Individual station reports are received too irregularly to permit even selected stations to be used for computation points.

Nevertheless, all available data are used each day. Analysis of each parameter is made by the continuous weighted-mean technique. Values are assigned from the technique to provide the necessary daily grid-point values from which soil-moisture values are computed. A rectangular grid, with grid points spaced approximately 60 nm apart (1/3 GWC grid spacing), is used.

(2) When insufficient input data are available to satisfy the analysis program for a grid point, a missing indicator is printed in lieu of a value. The analyst has the option of inserting estimated daily values of missing parameters if he desires. If reasonable estimates for missing daily values are not possible, the USAF ETAC computer program (soil moisture) inserts daily values based on long-term means when decade computations are made. Such missing daily grid-point values are rare in Europe but quite common in parts of Asia.

(3) The daily analysis of total precipitation for grid points is the most critical portion of the program, since the precipitation field is generally represented by a discontinuous type function. Monthly values of total precipitation are extracted from analyses for selected CLIMAT stations and verified against monthly totals reported by these stations. Verification results are generally good.

f. Changes in soil moisture during the decade are calculated by the following method:

(1) The potential evapotranspiration (PE) is calculated for each grid point by determining the yearly heat index (I) from Thornthwaite and Mather's Table 1.1 [48]. To do this, enter the table with the climatic monthly mean temperature for the particular grid point and extract the monthly heat index (i) for each month. The yearly heat index is the total of the 12 monthly i values. This procedure needs to be done only once, as each location or grid point and tables of I are in the prepared computer program. Next, determine the unadjusted PE daily value for each point from Figure 1.1 [48] by entering the figure with the appropriate mean 10-day temperature and the appropriate I value. Now, divide the unadjusted PE value by 28, 30, or 31, as applicable, to obtain the daily value. To obtain the adjusted PE, use Table 1.2, page 98 of the same reference. Enter the table with the appropriate latitude and month to extract the mean possible monthly duration of sunlight. Multiply the unadjusted PE value by this sunlight value to obtain an adjusted daily PE value. Multiply the adjusted daily PE value by ten to obtain the 10-day value. Note that the 10-day mean temperature is the only variable parameter involved in the computation of PE.

(2) P minus PE is then calculated for each grid point.

(3) Current soil moisture, as of the end of the decade, is determined

by adding the change in soil moisture (ΔSM) to the soil moisture at the end of the previous decade. The method of calculating ΔSM depends upon the sign of P minus PE, thus, if P minus PE is positive, then

$$\Delta SM = P \text{ minus } PE$$

if P minus PE is negative, then

$$\Delta SM = (P \text{ minus } PE) \times \frac{\text{Previous SM}}{200}$$

(4) For soil-moisture calculations, values in excess of 200 mm of water per top meter of soil are discarded at this point. However, tractionability class information is determined directly from soil-moisture information before the excess of 200 mm of water per top meter of soil is discarded. Descriptions of tractionability classes, which appear on pages 20 and 21 of Air Force Surveys in Geophysics No. 94 [50], apply to Table 28, except that a soil depth of one meter is used instead of the two feet noted in referenced descriptions. Table 28 gives the tractionability classes that appear, as numbers, on tractionability charts.

TABLE 28

Tractionability Classes.

No. on Charts	Class	Average Soil Moisture in One-Meter Depth (% of Field Capacity)	Tractionability	
			Plastic Soils	Sandy Soils
0	Frozen surface	Average mean temp. was less than 0°C	Improved if moist, little changed if dry	
1	Very dry	Less than 33%	Good	Poor to very poor
2	Dry	33-75%	Good	Poor
3	Moist	75-115%	Deteriorates rapidly in this range	Good
4	Very moist	115-155%	Poor	Excellent
5	Wet	155-200%	Nearly impossible	Fair

g. In the interpretation and use of tractionability charts, there are three important considerations:

(1) As previously stated, grid values must be considered as average values over an area; the averaging process depends on how much data are available in the vicinity of each particular grid point.

(2) Since soil moisture and tractionability calculations are made at the end of every decade (the last day of the month being the end of the third decade for the month), an assumption of an even distribution of precipitation throughout the decade is inherent in the computations. Thus, if most precipitation falls early in the decade, actual soil moisture will be somewhat less and the tractionability class number somewhat lower than that calculated. The converse is true if most precipitation falls late in the decade. Soil moisture and tractionability calculations are valid for the last day of the decade.

(3) Terrain effects on surface run-off and soil-moisture retention are not considered. Variations of the field capacity due to different soil types and root zones are not considered. Our assumption of a field capacity of 200 mm of water for the top meter layer of soil is reasonable for a clay loam soil with a grain crop; however, heavy clay soils hold more and sandy soils hold less water.

3. Climatological Wind Factors.

a. Climatological wind information is used in planning air operations when the planning period exceeds the time period for which the forecasting agency can supply reliable flight-level wind forecasts. In air route planning, the analyst is interested in the extent to which the wind either aids or retards the aircraft.

b. A wind factor can be defined as the difference between the ground speed of an aircraft and its true air speed. This relationship is given in Equation (146).

$$w = |G| - |A| \tag{146}$$

and is shown graphically in Figure 23, where G is the ground speed of aircraft, A the airspeed of aircraft, V the velocity of wind, w the wind factor, u the wind component along aircraft track, and v the crosswind component. The ground speed of the aircraft is given by

$$G = A \cos \alpha + u$$

$$= A(1 - \sin^2 \alpha)^{1/2} + u$$

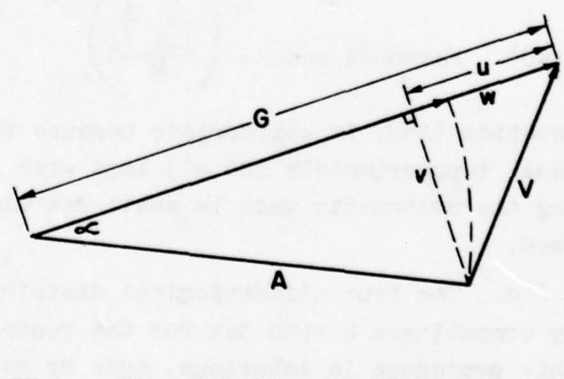


Figure 23. Relationship Between Ground Speed, Airspeed, Wind Velocity, and Wind Factor.

and, since $v = A \sin \alpha$,

$$G = A\left(1 - \frac{v^2}{A^2}\right)^{1/2} + u$$

From the binomial formula,

$$\left(1 - \frac{v^2}{A^2}\right)^{1/2} = 1 - \frac{1}{2}\frac{v^2}{A^2} + \frac{1}{8}\left(\frac{v^2}{A^2}\right)^2 - \frac{1}{16}\left(\frac{v^2}{A^2}\right)^3$$

Neglecting 4th powers and greater of v/A gives

$$\left(1 - \frac{v^2}{A^2}\right)^{1/2} \cong 1 - \frac{1}{2}\frac{v^2}{A^2}$$

so

$$G \cong A\left(1 - \frac{v^2}{2A^2}\right) + u$$

Therefore

$$w = G - A \cong A\left(1 - \frac{v^2}{2A^2}\right) + u - A$$

and the wind factor for a route is given by

(147) $$w \cong u - \frac{v^2}{2A}$$

c. The wind factor over a route = [w] = harmonic mean of w over all the legs of a route.

(148) $$\text{Harmonic mean} = \left(\frac{\sum_{i=1}^{k} \frac{1}{w_i}}{k}\right)^{-1}$$

Equation (148) is appropriate because the route wind factor is computed over equal (approximately 300 mi) legs with variable time. Since the error in using the arithmetic mean is small for wind speeds ≤ 1/3 A, arithmetic means are used.

d. The true climatological distribution of the wind factor (w) is found by computing a w each day for the route over an adequate period of record; this procedure is laborious, even by machine. If winds aloft are assumed circular normal, then the mean of w for a route may be estimated from V_r (resultant wind speed) and σ (standard vector deviation of wind velocity). The relationship between ground speed, airspeed, resultant wind direction, and

standard vector deviation of the wind are shown in Figure 24.

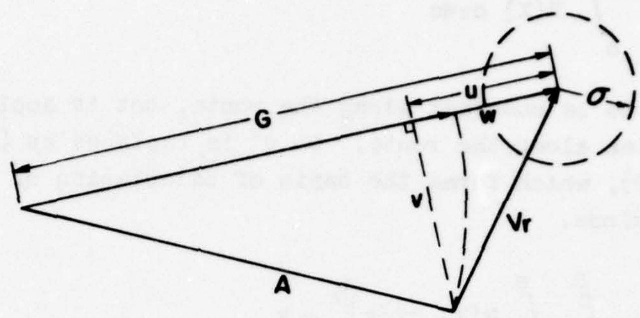


Figure 24. Relationship Between Ground Speed, Airspeed, Resultant Wind Speed, and Standard Vector Deviation of Wind Velocity.

From Equation (147), $w = u - v^2/2A$, and over a long period of time $\overline{w} = \overline{u - v^2/2A}$ and then $= \overline{u} - \overline{v^2}/2A$, where $\overline{v^2} = \overline{[\overline{v} + (v - \overline{v})]^2} = \overline{v}^2 + 2\,\overline{v}\,\overline{(v - \overline{v})} + \overline{(v - \overline{v})^2}$ and $\overline{(v - \overline{v})} = 0$ and $\overline{(v - \overline{v})^2} = \sigma_v^2$. Assuming a circular normal distribution, $\sigma_v^2 = (\sigma/\sqrt{2})^2$, so $\overline{v^2} = \overline{v}^2 + \sigma^2/2$ so $\overline{w} = \overline{u} - 1/2A\,(\overline{v}^2 + \sigma^2/2)$. This is a time (monthly) mean and, taking the equiva-tailwind for the route as the mean or average of all legs, gives empirically $[\overline{w}] = \overline{[w]}$. Therefore, the mean or climatological wind factor for a route is given by

$$(149) \quad [\overline{w}] = [\overline{u}] - \frac{1}{2A}\left([\overline{v}^2] + \frac{[\sigma^2]}{2}\right)$$

In Equation (149) the bar above the symbol indicates an average with time (e.g., monthly mean for a leg) and the [ ] indicates an average over all the legs. The first term is the tailwind component and the second term is the effect of the crosswind.

e. The standard deviation, σ_w, of the equivalent headwind over the route as a whole is given by

$$(150) \quad \sigma_w^2 = \sigma^2\left(1 + \frac{1}{4}\frac{\sigma^2}{A^2} + \frac{\overline{v}^2}{A^2}\right)\frac{1}{S^2}\int_0^S \int_0^s R(X)\, dx\, ds$$

where S is the route length, s a segment of total route, A the airspeed, v the crosswind component, and R(X) the correlation between effective tailwinds x miles apart along the route. The second and third terms in the parentheses

are usually negligible and the above equation reduces to

$$(151) \qquad \sigma_w^2 = \frac{\sigma^2}{S^2} \int_0^S \int_0^s R(X)\, dxds$$

where σ is assumed to be constant along the route, but is applicable to the problem when σ varies along the route. So σ^2 is replaced by $[\sigma^2]$ and this gives Equation (152), which forms the basis of calculation of the variability of equivalent headwinds.

$$(152) \qquad \frac{\sigma_w}{[\sigma^2]^{1/2}} = \left[\frac{1}{S^2} \int_0^S \int_0^s R(X)\, dxds\right]^{1/2} = k$$

Equation (152) has been evaluated by Sawyer [42] from values of the correlation coefficient, R(X), determined from two years' measurements of the geostrophic wind on the 500-mb surface over the North Atlantic. These values of the k-factor are tabulated in Table 29.

f. Percentiles of the wind factor may be found from

$$(153) \qquad W_{\%} = [\overline{W}] \pm Z\, \sigma_w$$

where Z is the standardized cumulative normal deviate. Percentiles are used when risk is involved; 10% risk, called "90% worst wind factor," is most commonly used. The 90% worst wind factor is

$$(154) \qquad W_{90\%} = [W] - 1.28\, \sigma_w$$

Example 45: What is the mean and 90% worst wind factor at 15,000 feet for a flight from Pittsburgh to Miami during November? Airspeed of aircraft: 250 K.

TABLE 29

Factor to Convert Mean Standard Vector Deviation of Winds Over a Route (σ) to Standard Deviation of the Wind Factor (σ_w).

Route Length (nm)	Factor $\left(k = \frac{\sigma_w}{[\sigma]}\right)$
0	.71
200	.69
400	.67
600	.65
800	.62
1000	.60
1200	.58
1400	.56
1600	.53
1800	.51
2000	.49
2200	.47
2400	.46
2600	.45
2800	.43
3000	.42
3200	.41
3400	.40
3600	.39
3800	.38
4000	.37

Step 1. Divide the track into legs (equal legs, if possible).

Route: Pittsburgh, Pennsylvania to Miami, Florida.

Airspeed: 250 K.

Track: 180°

Length: Approximately 1000 miles.

Legs: 1. Pittsburgh to 37°.

2. 37°N to 33°N.

3. 33°N to 29°N.

4. 29°N to Miami.

Step 2. Determine V_r and σ for each leg from mean charts, SAC Manual, summaries, or other source, and $[\sigma]$ and $[\sigma^2]$, where [ ] indicates an average for all the legs. The following values are obtained:

	V_r	σ	σ^2
Leg 1.	265/50	34	1156
Leg 2.	268/48	32	1024
Leg 3.	268/40	28	784
Leg 4.	230/30	24	576
			3540

then calculating

$$[\sigma^2] = \frac{\Sigma \sigma^2}{\text{No. of legs}} = \frac{3540}{4} = 885$$

and

$$[\sigma] = \sqrt{[\sigma^2]} = \sqrt{885} = 29.7$$

Step 3. Determine $\overline{u}$ (the component of V_r along the track, positive if helping wind), $\overline{v}$ (the crosswind component of V_r), and $\overline{v}^2$. Since

$$\overline{u} = V_r \text{ Cos } \theta \qquad \text{and} \qquad \overline{v} = V_r \text{ Sin } \theta$$

where θ is the angle between the track and the mean wind, the following data are obtained:

	V_r	θ	Cos θ	$\overline{u}$	Sin θ	$\overline{v}$	$\overline{v}^2$
Leg 1.	265/50	85	.087	-4.4	.996	49.8	2480
Leg 2.	268/48	88	.035	-1.7	.999	48.0	2304
Leg 3.	268/40	88	.035	-1.4	.999	40.0	1600
Leg 4.	270/30	90	.000	0.0	1.000	30.0	900
				-7.5			7284

Step 4. Determine $[\overline{u}]$ and $[\overline{v}^2]$ where

$$[\overline{u}] = \frac{\Sigma\,\overline{u}}{\text{No. of legs}} = \frac{-7.5}{4} \approx -1.9$$

and

$$[\overline{v}^2] = \frac{\Sigma\,\overline{v}^2}{\text{No. of legs}} = \frac{7284}{4} = 1821$$

Step 5. Determine the mean wind factor W from Equation (149):

$$W = [\overline{w}] = [\overline{u}] - \frac{1}{2A}\left([\overline{v}^2] + \frac{[\sigma^2]}{2}\right)$$

$$W = -1.9 - \frac{1}{2 \times 250}\left(1821 + \frac{885}{2}\right)$$

$$= -1.9 - \frac{1}{500}\,(1821 + 443)$$

$$= -1.9 - \frac{2264}{500} = -1.9 - 4.5$$

$$W = -6.4 \text{ knots}$$

Step 6. If the "90% worst," 10% risk wind factor, is desired, determine the standard deviation of wind factor σ_w from

$$\sigma_w = k\,[\sigma]$$

where k is the adjustment factor given in Table 29 for a route length of 1000 miles, k = 0.60, therefore

$$\sigma_w = 0.60 \times 29.7 = 17.8$$

Finally, determine the "90% worst" from

$$W_{90} = W - 1.28\,\sigma_w$$

where 1.28 is the .90 deviate from the cumulative normal distribution

$$W_{90} = -6.4 - 1.28 \times 17.8 = -6.4 - 22.8 = -29.2 \text{ knots}$$

4. Frequency of Occurrence of Nighttime Illuminations and Various Ceilings and Visibilities.

a. This technique involves finding the frequency of the joint occurrence of illumination values and various ceiling/visibility categories. The material used includes seasonal charts of nighttime illumination and seasonal graphs of ceiling/visibility categories.

b. Curves for the frequency of occurrence of nighttime illumination at 40 and 50 degrees north latitude were prepared with the aid of illumination values obtained by Brown [13] (see Figure 30). These values were obtained from more than 1200 illumination measurements made by Dayton R. E. Brown in the Arctic, Antarctic, and the temperate and torrid zones of both hemispheres between January 1943 and May 1947. The curves presented (Figures 25 through 28) are for clear sky conditions only. No attempt has been made to determine the amount of attenuation of nighttime illumination due to clouds. The values of illumination are in footcandles. A footcandle is defined as the luminous energy received on any part of a surface per unit of time when the surface is normal to and one foot distant from a light power source of one international candle. The illumination curves in this report are restricted to the nighttime values bounded by civil twilight. The lower limit of civil twilight occurs when the sun is six degrees below the horizon and the illumination is 3.16×10^{-1} footcandle.

c. The annual and seasonal ceiling/visibility values presented in Table 30 were obtained from information contained in USAF ETAC Report 4803 (Flying Weather in Central Europe) [8].

TABLE 30

Ceiling/Visibility Percentages for Central Europe (All-Hours)

	Jan	Apr	Jul	Oct	Annual
≥ 500 ft and ≥ 3 mi	58	87	88	71	76
≥ 2000 ft and ≥ 3 mi	42	73	78	59	63
≥ 2000 ft and ≥ 6 mi	26	61	65	42	48

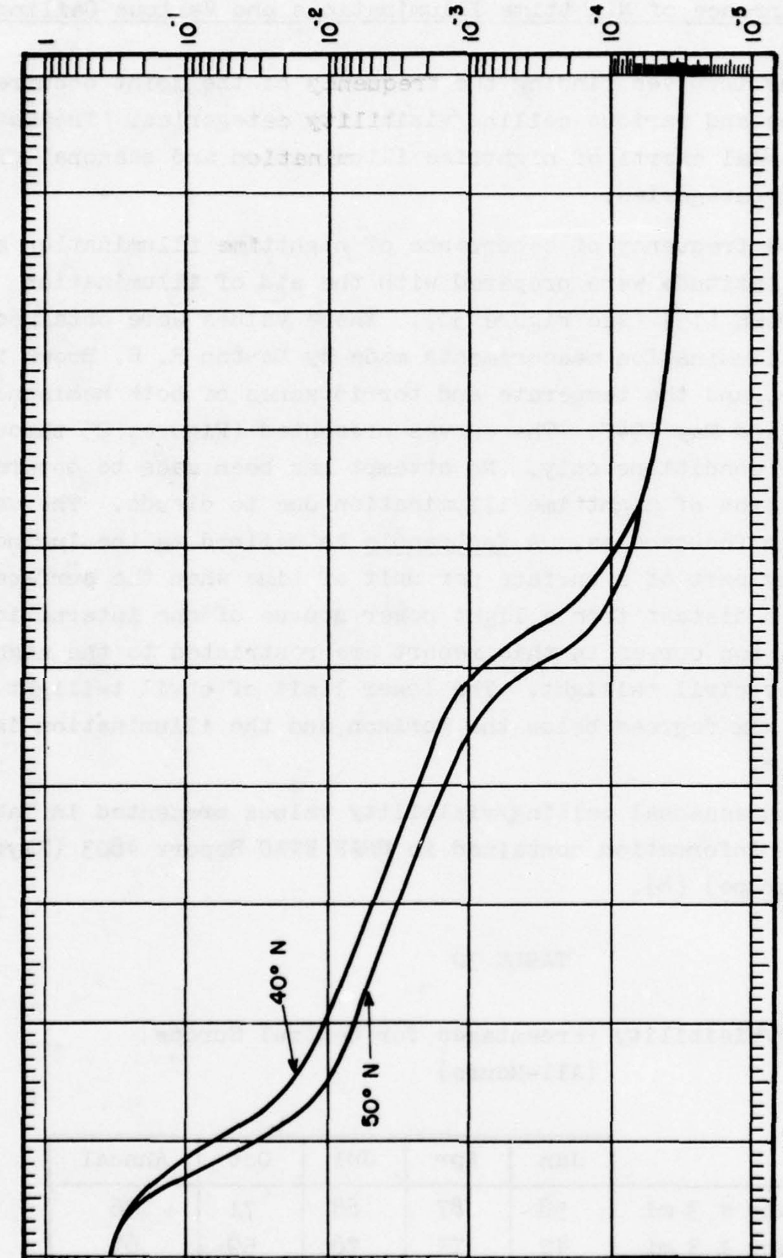


Figure 25. Percent Time Nighttime Illumination is Equalled or Exceeded in January for Northern Hemisphere.

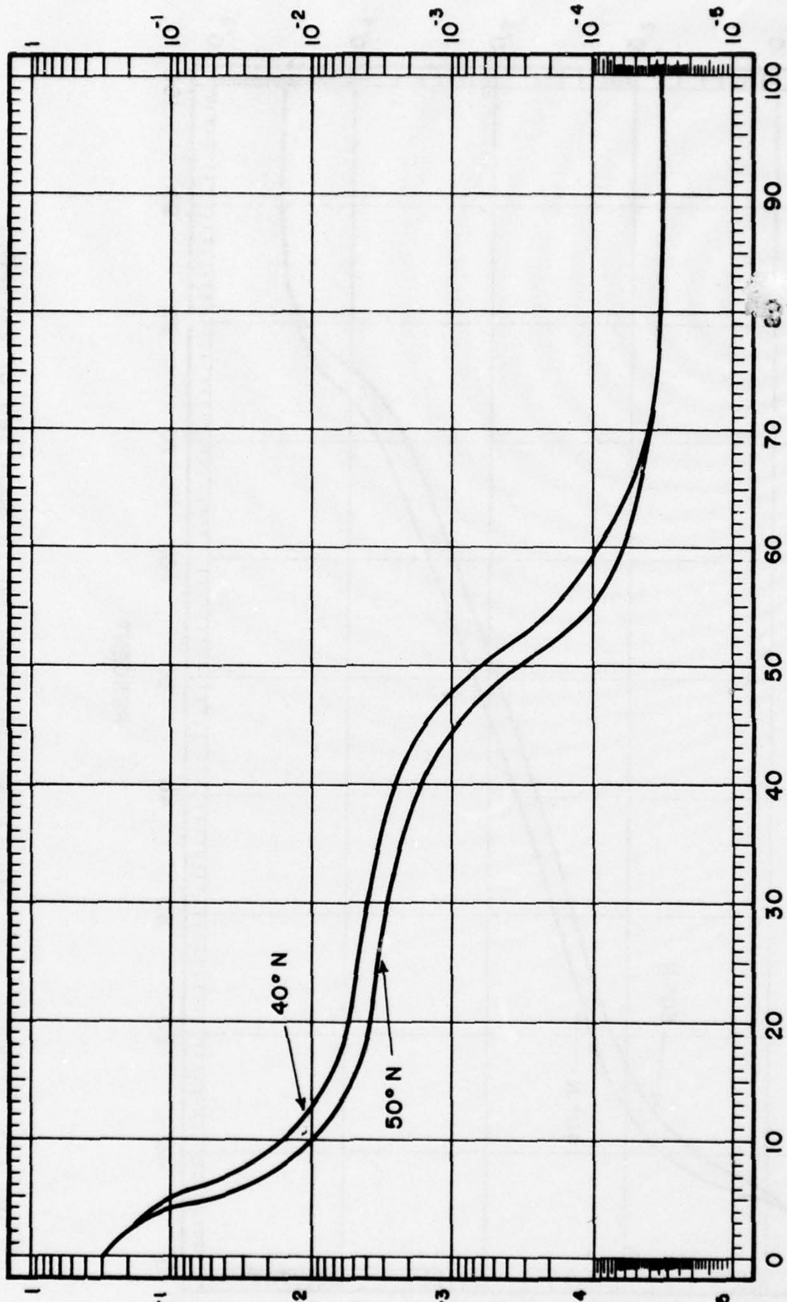


Figure 26. Percent Time Nighttime Illumination is Equalled or Exceeded in April for Northern Hemisphere.

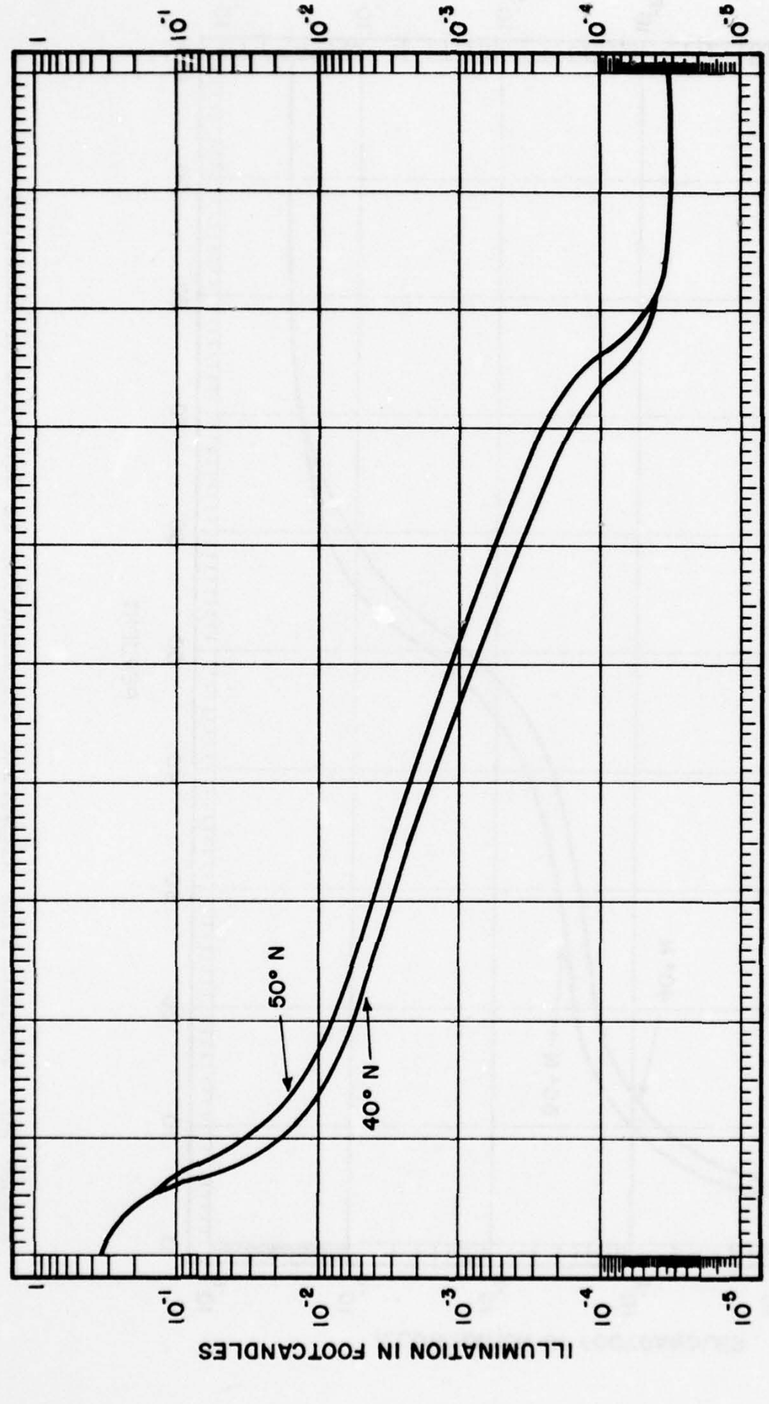


Figure 27. Percent Time Nighttime Illumination is Equalled or Exceeded in July for Northern Hemisphere.

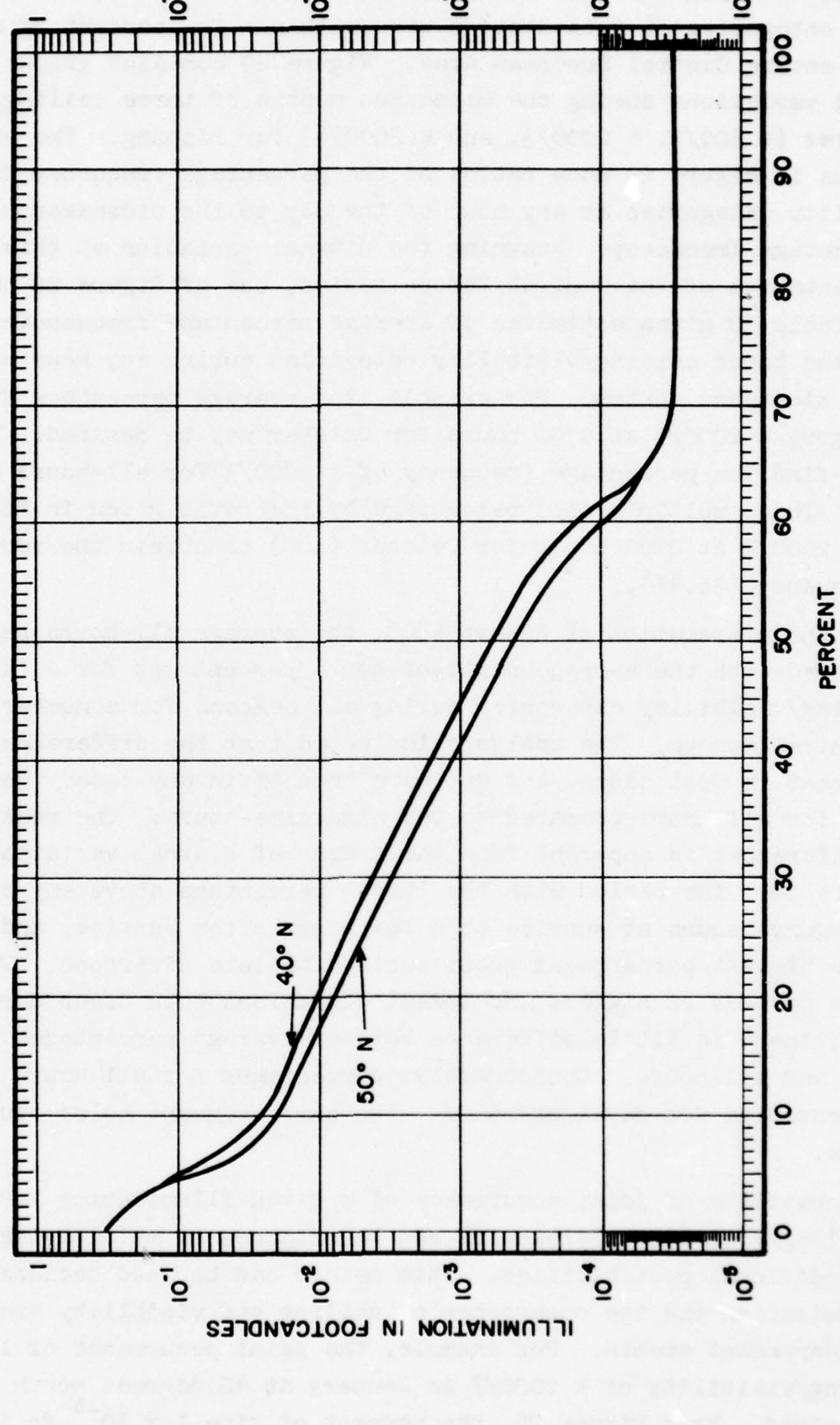


Figure 28. Percent Time Nighttime Illumination is Equalled or Exceeded in October for Northern Hemisphere.

d. The analyst studied the diurnal variation during the midseason months of various ceiling/visibility categories to determine a representative station in West Germany. Bitburg Air Base was selected because percentages for ceiling/visibility categories at this station were within a few percent of average values for the entire Central European Area. Figure 29 contains graphs representing diurnal variations during the midseason months of three ceiling/visibility categories (≥ 500/3, ≥ 2000/3, and ≥ 2000/6) for Bitburg. The ordinates for graphs in Figure 29 show ratios of the percentage frequency for the ceiling/visibility categories at any hour of the day to the midseason months all-hours percentage frequency. Assuming the diurnal variation at this station is representative of the Central Europe region, use of Figure 29 in combination with Table 30 gives estimates of average percentage frequencies of occurrence of the three ceiling/visibility categories during any hour or period for the midseason months. For example, the average percentage frequency of category ≥ 2000/3 at 0700 hours for October may be desired. First, from Table 30, find the percentage frequency of ≥ 2000/3 for all-hours in October (59%). Then, multiply this percentage by the ratio given in Figure 29 for category ≥ 2000/3 at 0700 hours for October (.60) to obtain the required percentage frequency (35.4%).

e. During the preparation of Report 4803, the average all-hours percentages were compared with the average daylight-hours percentages for a large number of ceiling/visibility categories during all seasons for a number of stations in Central Europe. The analysis indicated that the difference was only a few percent in most cases, and not more than 5% in any case. The same thing was true for all-hours compared to the nighttime-hours. The reason for these small differences is apparent from the graphs of diurnal variation presented in Figure 29. The period with the lowest percentage above any ceiling/visibility category occurs at sunrise or a few hours after sunrise, and the period with the highest percentages occur during the late afternoon. Therefore, since the periods of highest and lowest conditions both occur during daylight hours, there is little difference between average percentages for daylight-hours and all-hours. Consequently, percentages for all-hours may be used as representative for daylight-hours. The same argument holds true for nighttime-hours.

f. The probability of joint occurrence of a given illumination level and one of the ceiling/visibility categories can be obtained by multiplying together their individual probabilities. This method can be used because the nighttime illumination and the occurrence of ceiling and visibility are assumed to be independent events. For example, the joint occurrence of 1×10^{-4} fc and a ceiling/visibility of ≥ 2000/3 in January at 40 degrees north latitude may be desired. From Figure 25, the percent of time 1×10^{-4} fc is

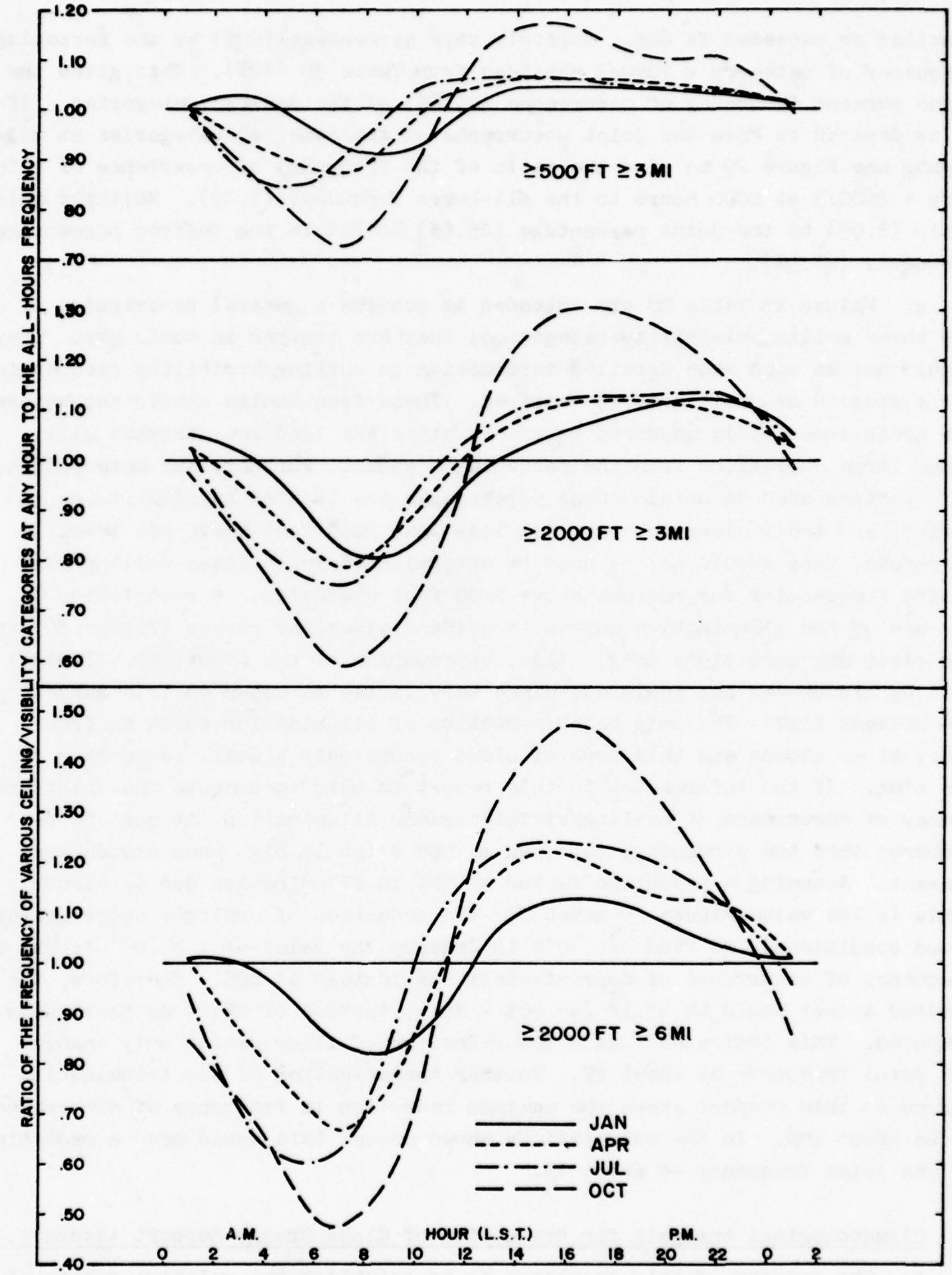


Figure 29. Diurnal Variation in the Occurrence of Various Ceiling/Visibility Categories — Bitburg AB, Germany.

equalled or exceeded is 60%. Multiply this percentage (60%) by the percentage frequency of category ≥ 2000/3 obtained from Table 30 (42%). This gives the joint percent frequency of occurrence (25.2%) of the desired categories. If it is desired to know the joint occurrence of the same two categories at midnight, use Figure 29 to find the ratio of the frequency of occurrence of category ≥ 2000/3 at 0000 hours to the all-hours frequency (1.09). Multiply this ratio (1.09) by the joint percentage (25.2%) to obtain the desired percentage frequency (27.5%).

g. Values in Table 30 are intended to provide a general description of the three ceiling/visibility categories; they are limited in their use. They should not be used when detailed information on ceiling/visibility frequencies for a station or small area is required. These frequencies should not be used for areas immediately adjacent to water, since the land-sea contrast will cause large variations from the percentages shown. Further, the meteorological stations used to obtain these percentages are located at airports or in cities, and their elevations are all less than 2000 feet above sea level; therefore, they should not be used to determine representative ceiling/visibility frequencies for regions above 2000 feet elevation. A restriction to the use of the illumination curves is evident since the curves (Figure 30) are for clear sky conditions only. Also, attenuation of the nighttime illumination by clouds was not included, since very little is known on this subject at the present time. The only major reduction of illumination would be from heavy storm clouds and this type of cloud occurs only a small percentage of the time. If the information in this report is used to compute the joint frequency of occurrence of ceiling/visibility and illumination, it must be remembered that the percentage obtained may be slightly high when clouds are present. Assuming a reduction factor of 50% in illumination due to clouds (this is the value Figure 30 gives for the reduction of sunlight under average cloud conditions), we find for 40°N in January the value of 1×10^{-4} fc has a frequency of occurrence of approximately 55% instead of 60%. Therefore, the desired answer would be 23.1% (or 55% × 42%), instead of 25.2% as previously computed. This indicates that a 50% reduction of illumination only changes the joint frequency by about 2%. Further investigation of the illumination curves in this respect shows the maximum reduction in frequency of occurrence to be about 10%. In the calculations shown above, this would mean a reduction in the joint frequency of about 4%.

5. Climatological Analysis for Evaluation of Close Ground Support Aircraft.

a. The purpose of this technique is to establish the relative merits of high and low performance aircraft in close ground support missions, based on their inability to operate below specified minimums of ceiling and visibility.

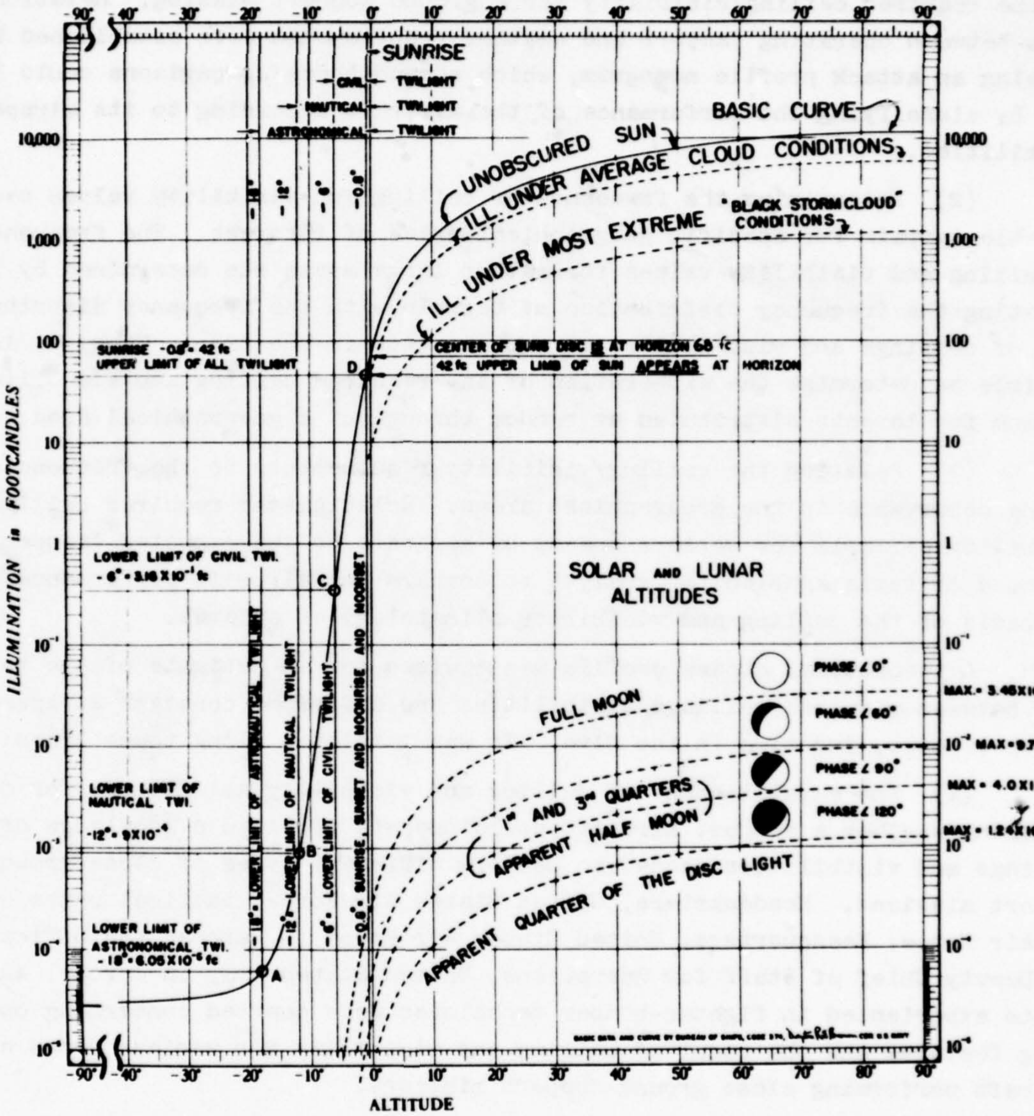


Figure 30. Total Range of Natural Illumination Levels.

There are three steps involved:

(1) Establishing relationships between specified operating factors and the required ceiling/visibility for a ground support mission. Relationships between operating factors and weather requirements were established by devising an attack profile nomogram, which showed these comparisons could be made by classifying the performance of the aircraft according to its airspeed capabilities.

(2) Determining the frequency of ceiling and visibility values over variable terrain for specific geographical areas of interest. The frequency of ceiling and visibility values for rather large areas was determined by integrating the frequency distribution of terrain with the frequency distribution of ceilings and visibilities at known points in the area. Thus, it is possible to determine the expectation of any required ceiling and visibility minimum for targets distributed at random throughout a geographical area.

(3) Relating the ceiling/visibility requirements to the frequencies of the occurrence in the geographical areas. Relating the required ceiling/visibility criteria for various speeds of aircraft to the expected frequency of those criteria enabled the analyst to compare the aircraft performance on the basis of the ceiling and visibility climatology of an area.

b. A theoretical attack profile was devised for an estimate of the relation between required ceilings/visibilities and different constant airspeeds, angles of dive, and time in the dive. It was initiated along these lines:

(1) The establishment of ceiling and visibility as criteria for determining whether a support aircraft could operate required a knowledge of the ceilings and visibilities needed to perform different types of close ground-support missions. Headquarters, United States Air Force; tactical units of the Air Force; Headquarters, United States Air Force in Europe; the Office of the Deputy Chief of Staff for Operations, United States Army in Europe; and pilots experienced in fighter-bomber techniques were queried concerning operating features and the required ceiling and visibility for various types of aircraft performing close ground-support missions.

(2) Hq USAF estimated the F-104 minimum ceiling and visibility to be 1500 feet and 3 miles within a cruise speed of 450 knots and a minimum speed of 285 knots. They estimated the Mohawk would require 500 feet and 1 mile with a cruise speed of 200 knots. Air Force tactical units estimated 6000- to 10,000-foot ceilings would be required for dive bombing, and a minimum of 500 feet and 3 miles would be required for strafing flat terrain (high performance aircraft implied). Hq USAF suggested a minimum acceptable flight visibility of three miles and a desirable visibility of five or more miles and the following ceilings for high performance aircraft to conduct:

(a) Low-level bombing (napalm or skip) 2500 feet with 1000 feet as a combat minimum.

(b) High-angle dive bombing — 10,000 feet.

(c) Rocketry — 6000 feet.

(d) Strafing — 2000 feet.

(e) GAM-83A — 20,000 feet.

(3) Hq USAEUR is quoted for minimum criteria of ceiling and visibility for low performance aircraft during field training exercises: fixed-wing aircraft, free of physical contact with clouds and visibility 1 mile or greater; and rotary-wing aircraft, free of physical contact with clouds and visibility 1/2 mile or greater.

(4) Pilots pointed out the following features: time spent in the actual dive should be on the order of 15 seconds; vertical clearance from the cloud deck should be at least 500 feet; clearance above the ground should be at least 100 feet, but depends on speed, angle of dive, and type of mission; and visibility depends on ceiling, speed of the aircraft, and terrain.

(5) Figure 31 shows the attack profile. The support aircraft approaches the target area in level flight, the target is located, the craft is put into its attack run, the attack is accomplished, and the aircraft departs the area. Several variables can be put into equations that can be solved for the ceiling and visibility required for the maneuver. Variables and the symbols representing them are:

θ = angle of dive. Angles of 5, 10, 15, 20, and 40 degrees were used in calculations to cover most of the practical range of dive angles.

A = airspeed in knots. Airspeeds are considered constant during the attack. Speeds ranging from 100 through 1000 knots were used.

a = acceleration (ft/sec^2).

a_1 = acceleration during pullout (ft/sec^2).

a_2 = acceleration (neg.) entering dive (ft/sec^2).

C_n = required minimum ceiling height for the maneuver (feet). The aircraft is assumed to be just clear at the start of the attack.

G = ratio of acceleration force to acceleration of gravity.

h_o = terrain clearance in pullout (feet).

h_1 = vertical distance from point of alignment to start of recovery (feet).

R = radius of curvature (feet).

S = airspeed (feet/second).

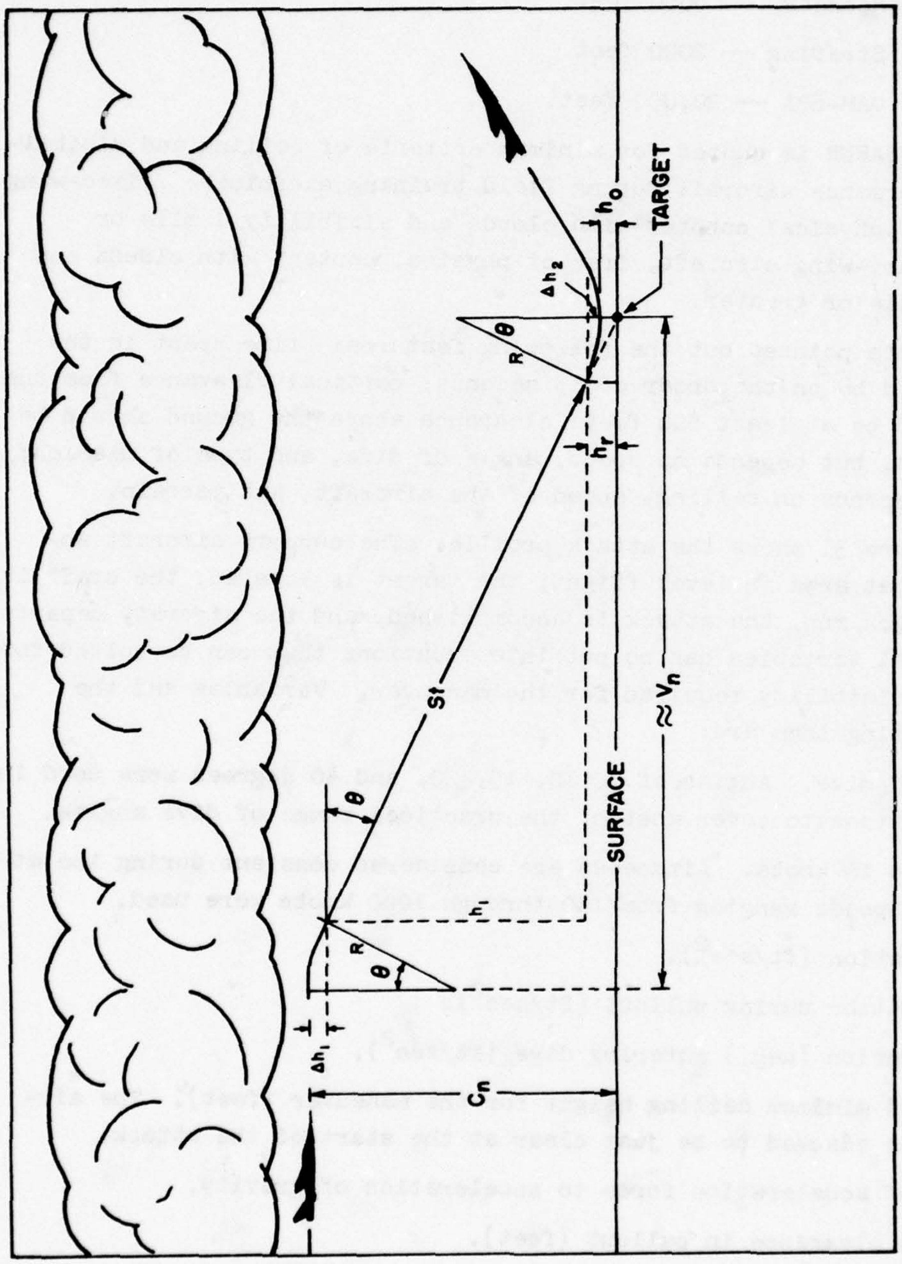


Figure 31. Theoretical Attack Profile.

t = time from point of alignment to start of recovery (seconds).

V_n = minimum visibility from start of attack to recovery (miles).

Development of the equations follows:

General equation for ceiling:

$$a = S^2/R$$

$$\Delta h = \Delta h_1 = \Delta h_2 = R - R \cos\theta = (S^2/a)(1 - \cos\theta)$$

$$h_r = h_o + h = h_o + (S^2/a)(1 - \cos\theta)$$

$$C_n = \Delta h + h_1 + h_r$$

$$C_n = (S^2/a_1)(1 - \cos\theta) + S\,t \sin\theta + h_o + (S^2/a_2)(1 - \cos\theta)$$

Let $a_1 = a_2 = a$

$$C_n = S\,[t \sin\theta + (2S/a)(1 - \cos\theta)] + h_o$$

Conversion to common units:

$$C_n = h_o + 1.69\,A\,[t \sin\theta + .105\,(A/G)(1 - \cos\theta)]$$

General equation for V_n (slant visibility):

$$V_n \approx V_{horizontal}$$

$$V_n \approx S\,t + 2R \tan\theta\, S\,t + 2(S^2/a) \tan\theta\, S\,t + (2S/a) \tan\theta$$

Conversion to common units:

$V_n \approx .00032\,a\,[t + (.105A/G) \tan\theta]$. The G term is defined by letting $G = 1 + (A/200)$, where A/200 is a ratio allowing for variable G limits in relation to the speed of the aircraft; a higher speed (higher performance aircraft) pulls more G forces. Then for

A = 150	200	400	600	1000	knots
G = 1.75	2.00	3.00	4.00	6.00	

(6) Figures 32, 33, and 34 are nomograms to determine the height loss and the minimum required visibility for a theoretical attack profile with t = 5, 10, and 15 seconds, respectively. Terrain clearance is eliminated to maintain the versatility of the nomograms. The required ceiling is obtained by adding the desired terrain clearance (h_o) to the height value obtained from the nomogram and, if applicable, required initial clearance between the aircraft and cloud base.

(7) The V_n value is the horizontal distance required for the maneuver,

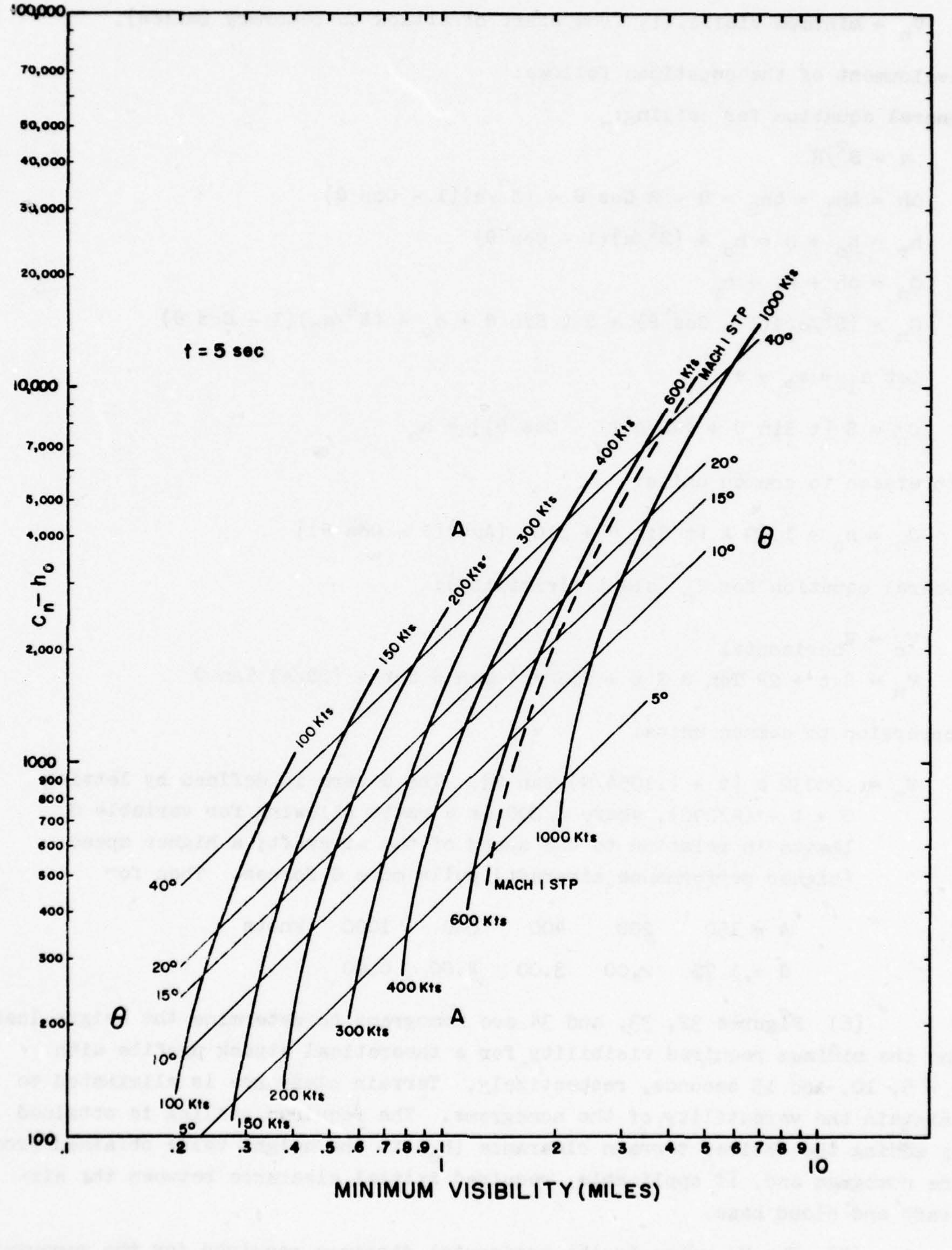


Figure 32. Attack Profile Nomogram, t = 5 sec.

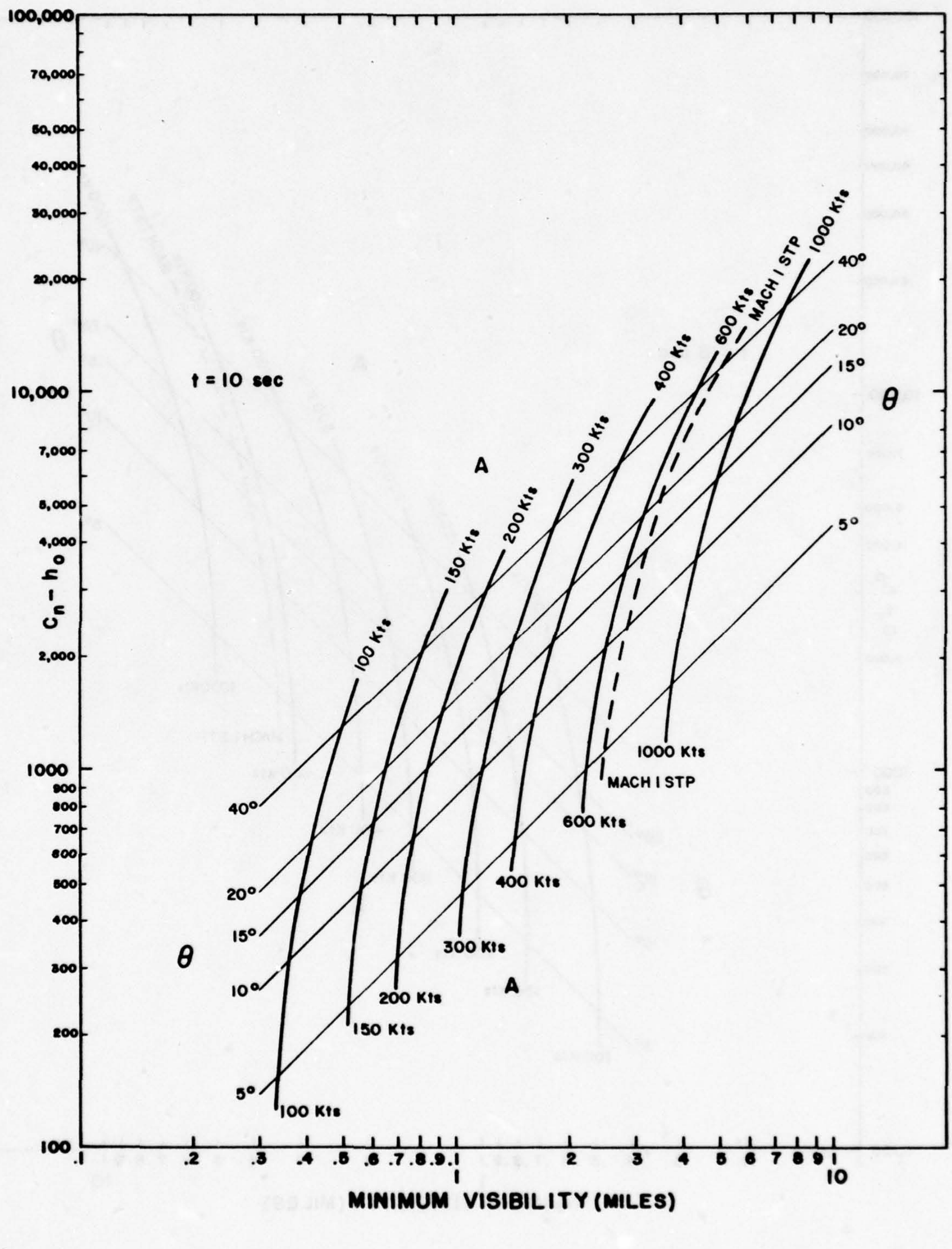


Figure 33. Attack Profile Nomogram, t = 10 sec.

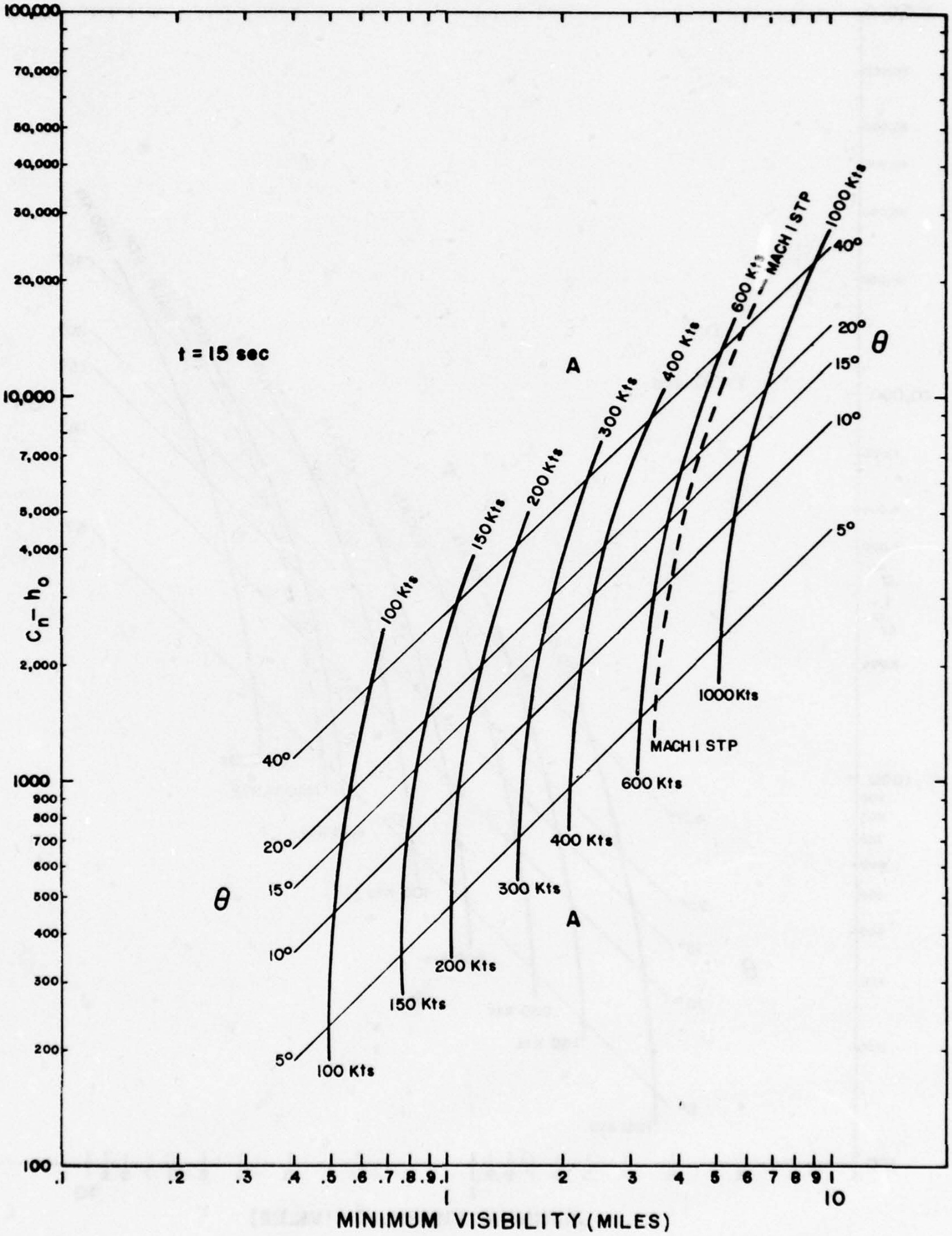


Figure 34. Attack Profile Nomogram, t = 15 sec.

therefore, a minimum visibility requirement. Most likely the pilot would need to see the target some time before he starts the dive. If the values of V_n given by the nomogram are multiplied by a factor of 1.5, the visibility requirement has a value closer to those suggested by the Air Force sources previously queried. The nomograms were used to establish ranges of ceilings and visibilities needed in climatological analyses of specific geographical areas.

c. A terrain versus ceiling and visibility analysis is required because ceilings are measured from ground to cloud base and ceiling heights over an area depend in part on terrain elevation.

(1) This terrain-height effect must be incorporated in the analysis of ceiling-height frequencies over an area. The procedure is to convert measured ceiling heights to the equivalent altitude of cloud bases above sea level; then ceilings above other points in the vicinity are determined by subtracting terrain elevations at these points. This assumes that cloud deck bases are substantially level. Examination of data for stations at different elevations indicates the assumption is valid.

(2) Contour maps for Europe, Southeast Asia, and Korea were used to determine the frequency distribution of terrain heights (Figure 35). By a process of weighting and integration, the analyst was able to apply ceiling and visibility data from available weather observing sites to the terrain-height distribution of the area being studied to determine ceiling-height frequencies over the area.

(3) Climatological data of selected stations throughout the geographical areas of interest were analyzed. The theoretical attack profile nomograms indicated an analysis of ceilings from approximately 100 through 25,000 feet and visibilities of 1, 3, and 6 miles would be adequate for the study. Ceiling/visibility data were analyzed for annual frequencies and "worst season" frequencies when one could be determined. Results are shown in Figures 36 through 40. Each of the figures shows the frequency of ceiling and/or visibility being less than the specified values. The selection of "worst season," in some cases, is rather arbitrary. For example, in Korea the poorer visibilities occur in winter while the frequency of ceilings is greater in summer.

d. The analyst compared a 200-, 400-, and 600-knot required speed for a hypothetical strafing mission. Each aircraft required a 10-degree dive for 10 seconds and terrain clearance of 100 feet to accomplish the mission. Referring to Figure 31, the nomogram furnished the C_n-h_o and V_n values. The required ceiling, C_n, is determined by adding the desired terrain clearance. Values for each speed are:

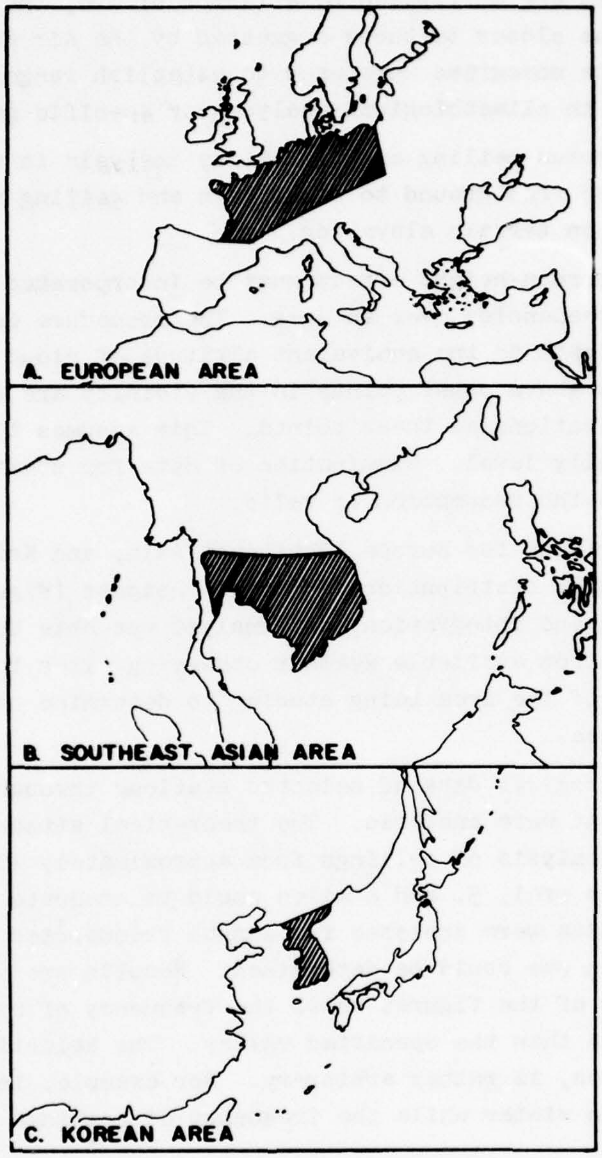


Figure 35. Areas Represented by Ceiling/ Visibility Study.

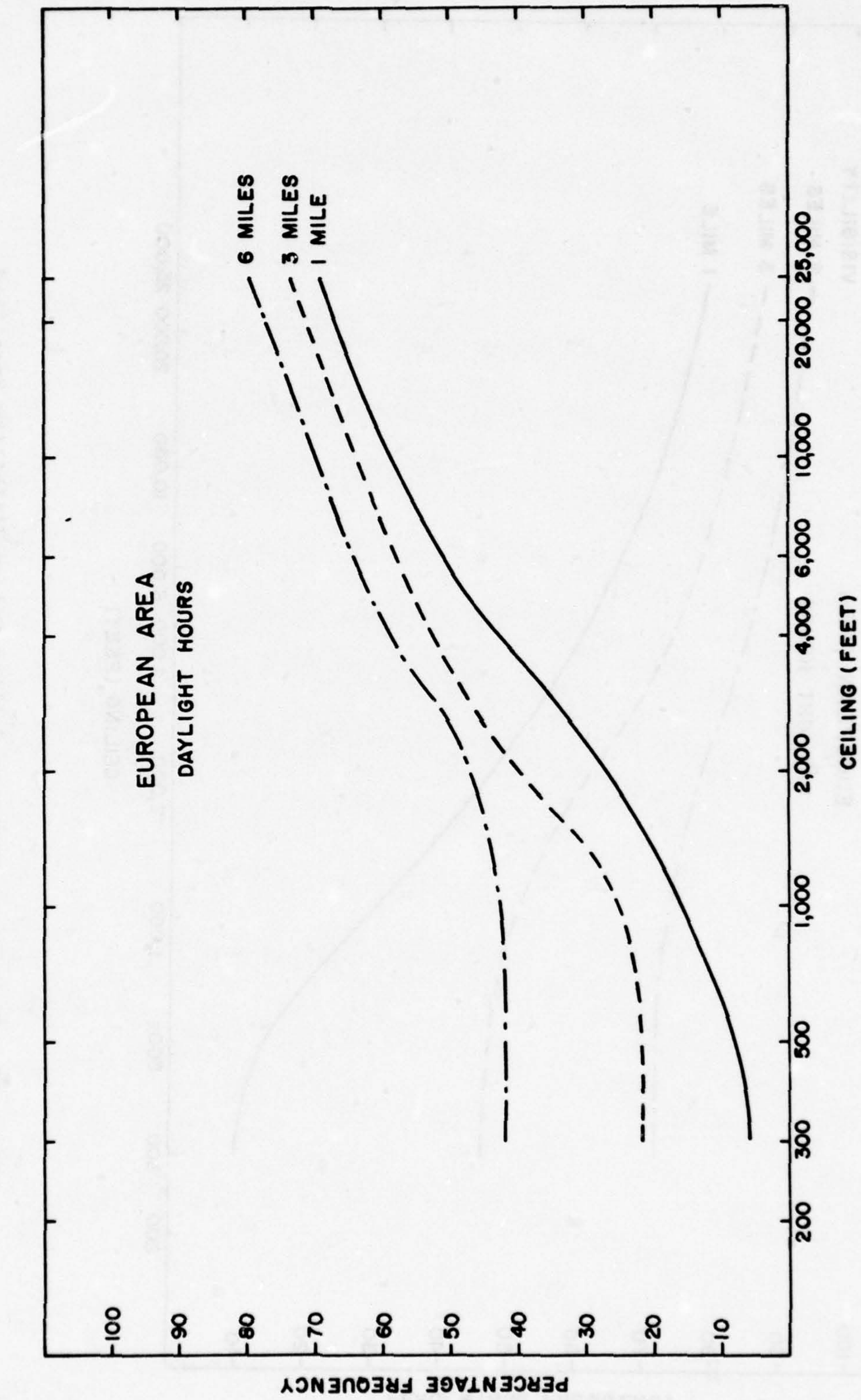


Figure 36. Annual Percentage Frequency Ceiling and/or Visibility Less Than Specified Values — European Area.

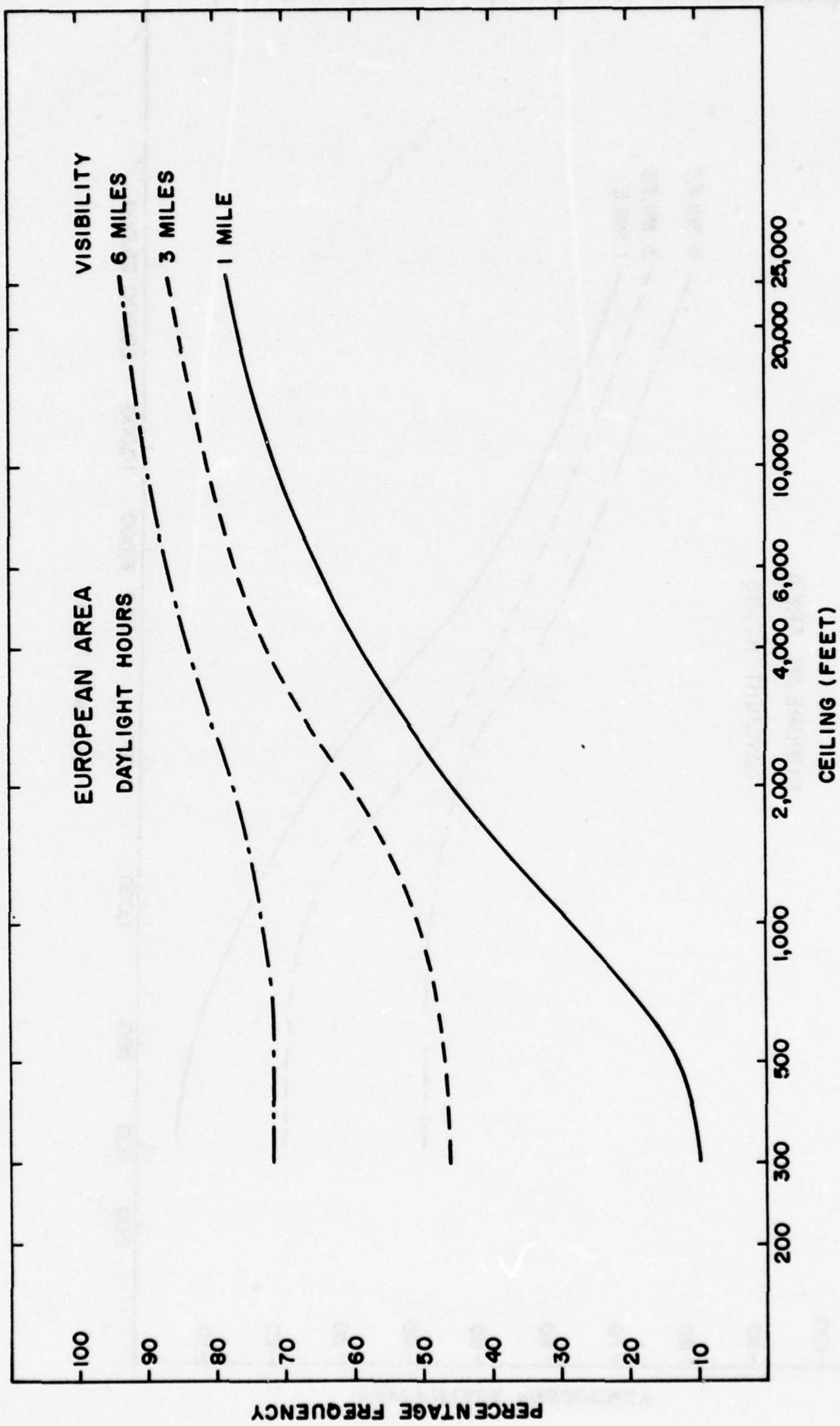


Figure 37. Percentage Frequency Ceiling and/or Visibility Specified Values — European Area.

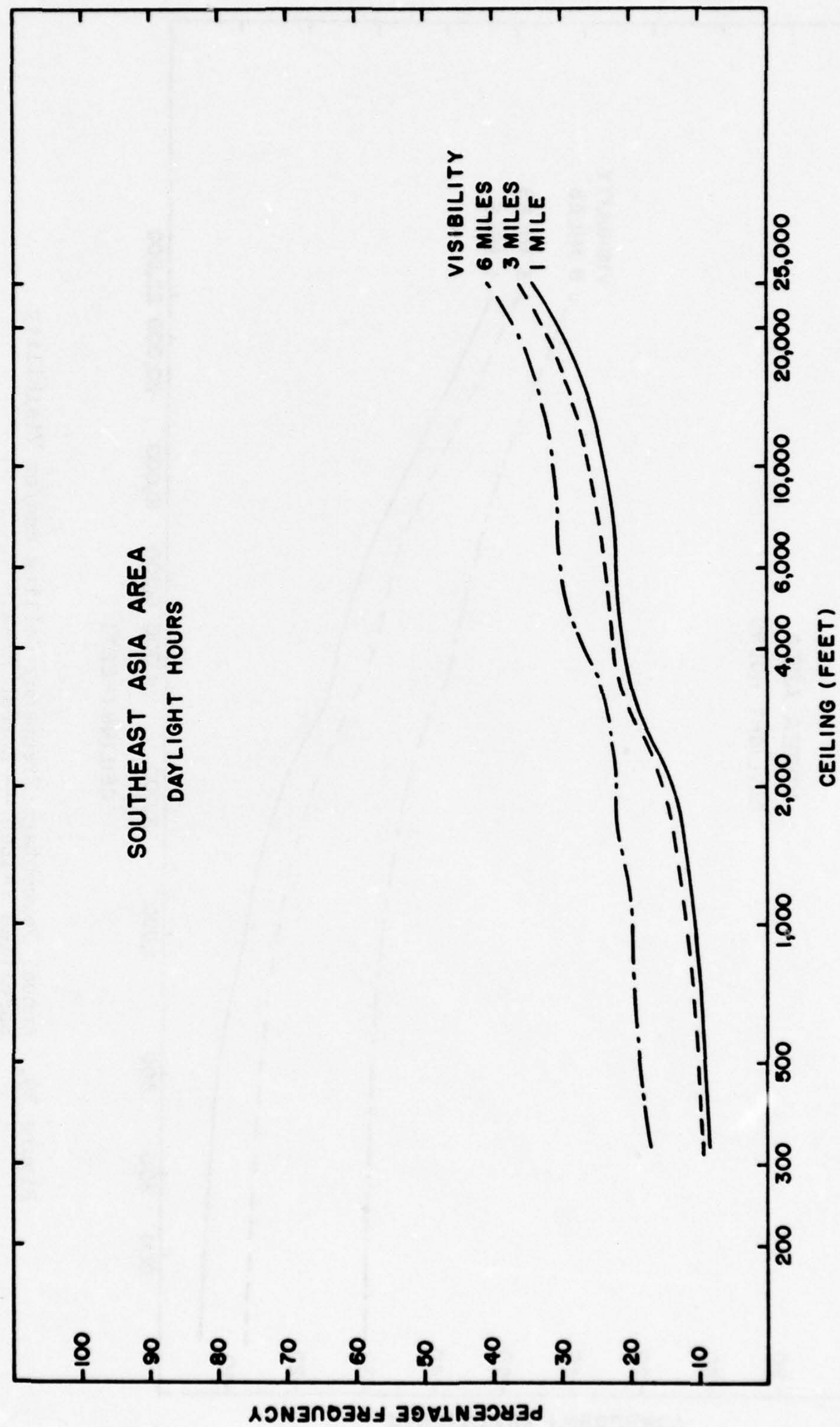


Figure 38. Annual Percentage Frequency Ceiling and/or Visibility Specified Values — Southeast Asia Area.

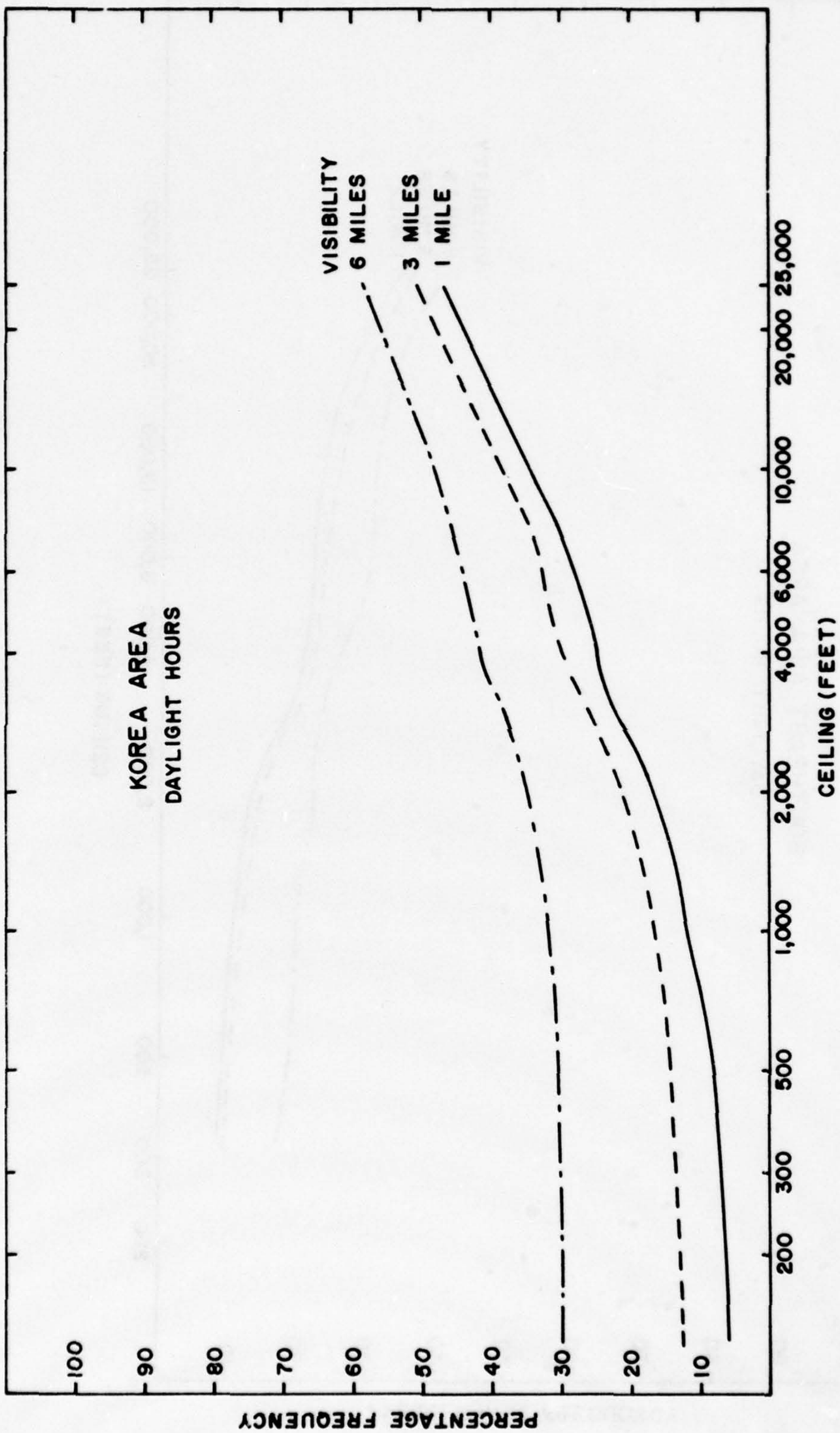


Figure 39. Annual Percentage Frequency Ceiling and/or Visibility Specified Values — Korean Area.

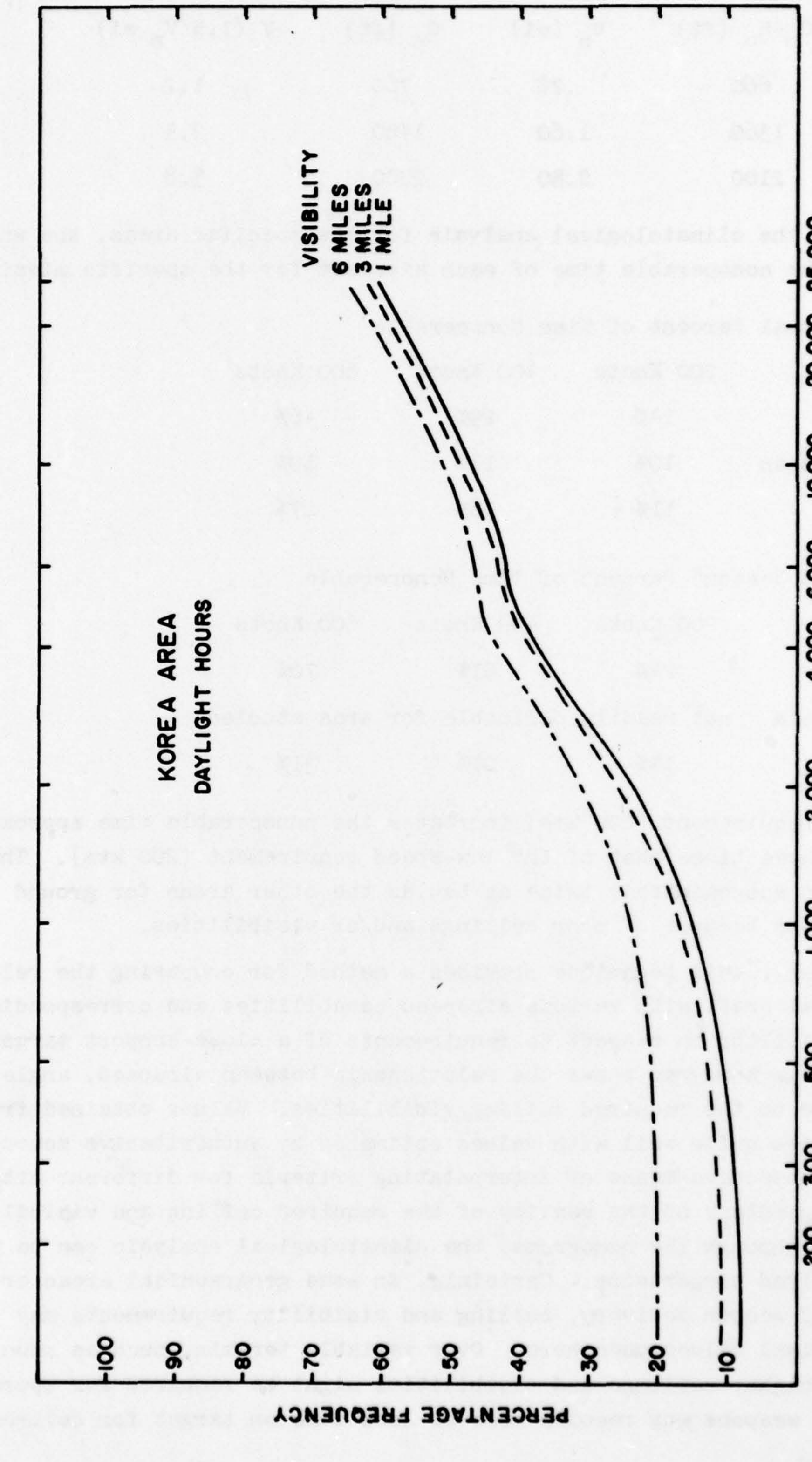


Figure 40. Summer Percentage Frequency Ceiling and/or Visibility Specified Values — Korean Area.

A (kt)	$C_n - h_o$ (ft)	V_n (mi)	C_n (ft)	V (1.5 V_n mi)
200	660	.78	760	1.2
400	1360	1.60	1460	2.4
600	2100	2.50	2200	3.8

By referring to the climatological analysis for the specific areas, the analyst compared the nonoperable time of each aircraft for the specific mission.

Annual Percent of Time Nonoperable

	200 Knots	400 Knots	600 Knots
European	14%	29%	46%
Southeast Asian	10%	13%	19%
Korean	11%	18%	27%

"Worst Season" Percent of Time Nonoperable

	200 Knots	400 Knots	600 Knots
European	24%	51%	70%
Southeast Asia	not readily definable for area studied		
Korean	14%	21%	31%

The high-speed requirement (600 kts) increases the nonoperable time approximately two to three times that of the low-speed requirement (200 kts). The European area is approximately twice as bad as the other areas for ground support operations because of poor ceilings and/or visibilities.

e. In summary, this technique provides a method for comparing the relative merits of aircraft with various airspeed capabilities and corresponding ceiling and visibility in respect to requirements of a close-support target. The attack profile nomogram shows the relationship between airspeed, angle, and time of dive to the required ceiling/visibilities. Values obtained from the nomogram agree quite well with values estimated by authoritative sources, and furnish an objective means of interpolating criteria for different attack conditions. Regardless of the reality of the required ceiling and visibility values obtained through the nomograms, the climatological analysis can be used to make the desired comparisons. Certainly, in some geographical areas or for certain types of weapon delivery, ceiling and visibility requirements may differ from minimal values used here. Over variable terrain, such as mountainous areas, higher ceilings and visibilities might be required for operations. Special weapons may require more or less time on target for delivery.

There is no implication that a 600-knot fighter must always perform a ground-support mission at 600 knots; many of the "Century Series" can operate at lower speeds. Comparison between 200 knots and 400 knots shows less differences in the example as presented. As differences in speeds become less, differences in nonoperable time become insignificant.

6. Climatology of Atlantic Tropical Storms.

a. Tropical storms of the Atlantic and Caribbean have been the subject of much research, primarily aimed at improving the forecast of storm movement; many theoretical and empirical studies have been advanced to improve these predictions. Techniques (statistical, objective, subjective, and numerical) are complicated and most require a considerable amount of synoptic information concerning the nature of the circulation in and around the storm region. Even though new methods have been advanced and some achievements made in this respect, no outstanding successful prediction system has been developed. When pertinent information is not available for making a forecast, or when the centrally prepared forecast is unavailable to the forecaster, a probabilistic forecast utilizing a climatological approach may be desirable. In some cases a statement of probability concerning storm movement may be entirely sufficient for the user.

b. There are available many studies of various climatological aspects of tropical storms. However, few are designed to furnish a probabilistic forecast as the end product. Studies by Malone [35] and Appleman [10] are particularly interesting because they utilize the uncertainty of synoptic and statistical prediction schemes and treat these errors as vector quantities. However, these studies estimate the probability of hurricane movement based upon the short-period forecast movement.

c. In 1960, McCabe [33] prepared a series of preliminary reports on the climatology of Far East typhoons. In these reports he describes the adaptation of the circular normal distribution of vector quantities to the probability of a typhoon trajectory approaching a target from a known storm position. The results of "goodness of fit" tests seem to justify using this method, as developed, in spite of the fact that some of the statistics show distributions of storm movement vectors that are elliptical rather than circular.

d. If the McCabe method is considered applicable for use with Atlantic hurricanes, the following actions by an analyst are necessary:

(1) Data in punched-card form on Atlantic storm positions, with estimates of storm intensity, are obtained from the National Hurricane Research Project. The card deck is composed of approximately 10,000 cards covering the period 1896-1959. It consists of 0000Z and 1200Z storm position reports.

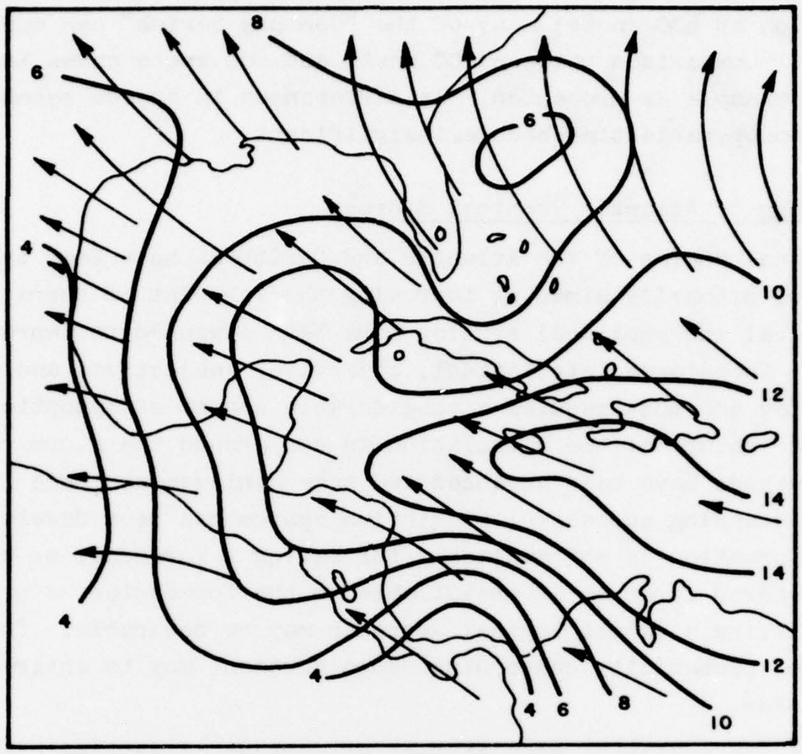


Figure 41. Resultant Vector Direction and Speed (kts).

(2) Assumptions concerning storm intensity classification and storm movement are made.

(3) The data are then entered on tape and the resultant vector direction (ψ) and speed (V_R), standard vector deviation of storm velocity (σv), and ratio of the standard vector deviation of storm velocity to the mean vector speed ($\sigma v/V_R$), among other parameters of storm movement, are computed for each 5-degree square.

(4) Pertinent values are then plotted on charts and an isoline analysis of each is prepared.

e. Charts of resultant vector direction and speed (Figure 41) and the ratio ($\sigma v/V_R$) (Figure 42) are used as basic work charts for computing composite trajectory frequency plots. One other basic chart, or set of charts, is required — the standard plot of the $\sigma v/V_R$ ratio as shown in Chapter 4, Section D of the CMF [9]. The standard plot must be adjusted for the particular target radius desired and the scale of the base map being used (Figure 43).

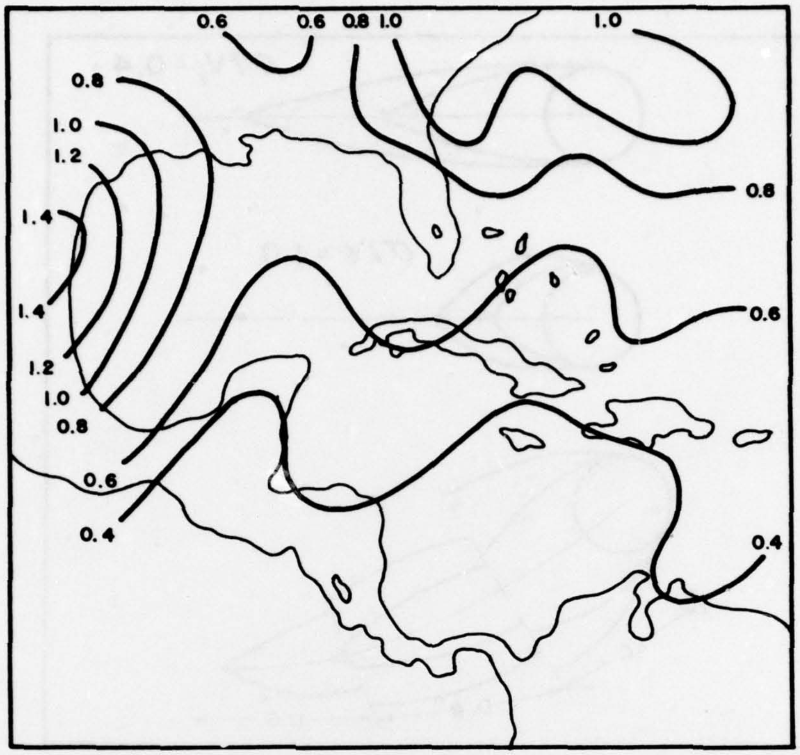


Figure 42. Ratio of Standard Vector Deviation to Mean Vector Speed (σ/V_R).

In studies of Atlantic tropical storms, the assumed storm or target radius is 60 nautical miles; in the Far East typhoon study, the assumed radius is 120 nautical miles.

f. The shape of the standard plot, as described in CMF 4-F [9], is determined by the $\sigma v/V_R$ value. Although these plots are intended to be applied where the descriptive parameters ($V_R, \sigma v$) are uniform, segments of the applicable plots can be combined to form a plot describing the probabilities of vectors reaching a target area. This is done in the following manner:

(1) From the charts of ψ and V_R, and $\sigma v/V_R$, indicate the upstream resultant vector (from the target) and the changes of the ratio along the upstream vector (Figure 43). This vector corresponds to the streamline through the point of interest.

(2) Trace sections of the applicable standard plots along the appropriate segment of the resultant vector. Smooth into a composite frequency plot.

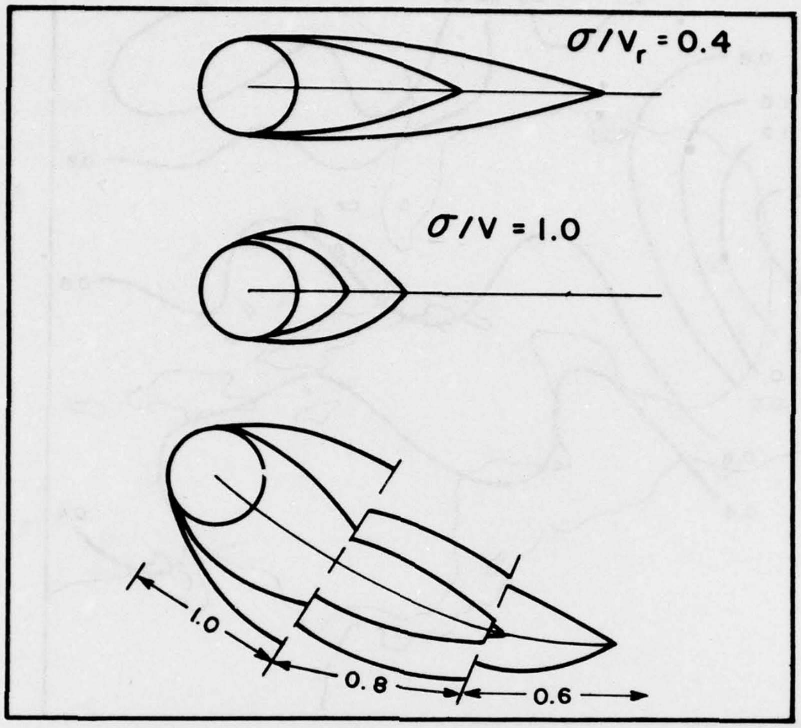


Figure 43. Standard Trajectory Frequency Plots.

(3) To give some indication of the distance a storm may move in a given time interval, a median 24-hour and 48-hour storm movement line can be drawn. The distance is determined by approximating the average V_R over the appropriate upstream distance. This distance is then an approximation to the median distance of storm movement.

g. These composite frequency plots provide a climatological estimate of the frequency that tropical storms and hurricanes (at random by reason of their present location) subsequently come within 60 nautical miles of the base location (Figure 44). That is, if a storm is located somewhere between the 10% and 20% isolines, the climatological probability of that storm coming within 60 nautical miles of the base is between .10 and .20; if the storm is located in the area delineated by the 50% and 60% isolines, the storm has a 50% to 60% chance of reaching the base area. Since plots are graphical descriptions of the normal behavior of storms during a period of time, a storm that is moving in opposition to the standard vector is probably less a threat than one following the normal track; however, no attempt has been made to assess

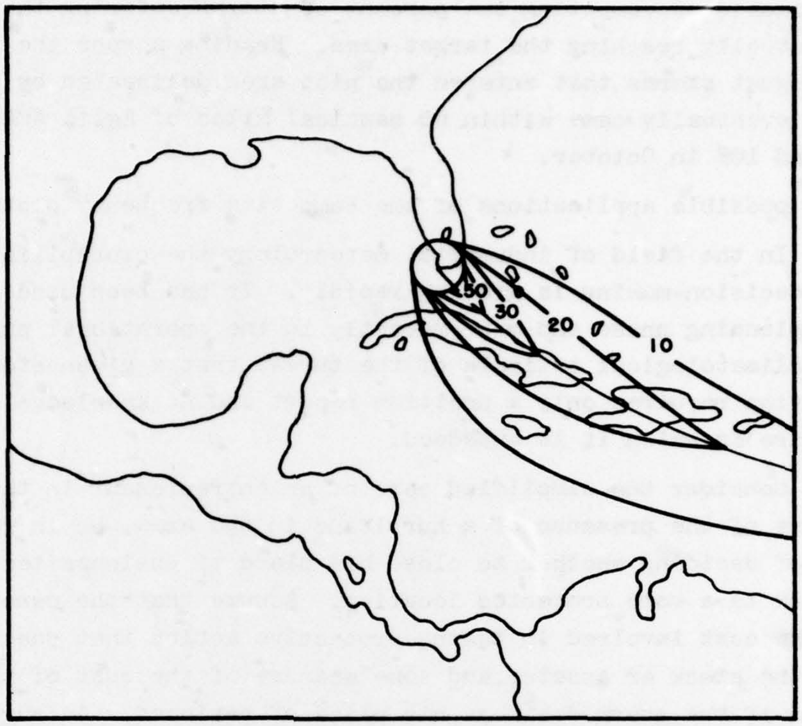


Figure 44. Composite Frequency Plot.

the short-period trajectory [7].

h. Tests of composite frequency plots with the storm tracks show that the plots are reasonable; however, a certain amount of subjectivity was inherent in the process of making a decision as to whether a storm was in or out of a particular zone. Since there are few storms to consider, a different subjective evaluation may alter the results in the high probability classes by 5 or 10%. The following are the results of this "goodness of fit" test on plots prepared for Eglin and Patrick Air Force Bases:

	Eglin			Patrick		
	Aug	Sep	Oct	Aug	Sep	Oct
10-20%	13	17	10	10	9	18
20-30%	27	36	20	19	25	35
30-50%	42	52	33	42	39	47
50-70%	73	60	45	77	70	69
> 70%	80	71	80	88	85	82

The test consisted of computing the percent of storms entering the indicated zone and eventually reaching the target area. Reading across the top line, 13% of the August storms that entered the plot area delineated by the 10% and 20% isolines eventually came within 60 nautical miles of Eglin AFB, 17% in September, and 10% in October.

i. Some possible applications of the composite frequency plots are:

(1) In the field of industrial meteorology the probabilistic approach to business decision-making is growing rapidly. It has been used for a long time in the planning phase and more recently in the operational phase. The use of this climatological estimate of the threat that a given storm poses to a given location requires only a position report and no knowledge of the complex atmosphere in which it is embedded.

(2) Consider the simplified case of an entrepreneur in the hurricane region. Aware of the presence of a hurricane in the area, he is faced with the problem of deciding whether to close his place of business temporarily and move his stock to a more protected location. Assume that the owner has some measure of the cost involved in taking protective action that does not involve movement of the stock or assets, and some measure of the cost of the alternative decision if the storm destroys his place of business. As suggested by Thompson and Brier [46], the following simple ratio may be a guide for decision:

$$P \gtreqless \frac{C}{L}$$

In this case, P is the probability of the storm coming within some prescribed distance, C is the total cost of taking protective action, and L is the potential loss. The decision criterion indicates that protective action should be taken when the probability that the hurricane will come into the area of interest is as great as or greater than the cost-loss ratio. Conversely, if the probability is less than the cost-loss ratio, no protective action need be taken. The economic factors involved in taking protective measures, estimating the uninsured loss, and insuring against prohibitive loss are complex, and a considerable amount of money is involved. In light of this, it may be of mutual interest for insurance companies to require protective measures at some level of storm threat. Such insurance may benefit both parties through reduced rates to the insured and a smaller loss of insured goods to the insurance company.

Chapter 7

REFERENCES AND BIBLIOGRAPHY

---

[1] Air Force: "Evaluation and Disposition of Records," AFM 181-5.

[2] Air Weather Service: "Acquisition of Data for Climatological Use," AWSR 105-3, 3 September 1965.

[3] Air Weather Service: "Applied Military Climatology," AWS Manual 105-3, 15 May 1968.

[4] Air Weather Service: "Project COLD FACT," AWSR 105-4, 29 April 1968.

[5] Air Weather Service: "An Aid for Using the Revised Uniform Summary of Weather Surface Observations," AWSP 105-1, 30 November 1966, pp. 2-14.

[6] Air Weather Service: "Climatic Briefs," AWSP 105-4, 20 December 1967, pp. 2-21.

[7] Air Weather Service, Climatic Center: "A Preliminary Report on the Climatology of Tropical Storms at Eglin and Patrick Air Force Bases," Climatic Center Report 3871, September 1961 (Unpublished).

[8] Air Weather Service, Climatic Center: "Flying Weather in Central Europe," Climatic Center Report 4803, February 1964.

[9] Air Weather Service, ETAC: "Climatic Methods File (CMF)," (Unpublished).

[10] Appleman, H. S.: "Estimating the Probability of Operationally-Critical Wind Speeds Affecting an Air Base during the Passage of a Tropical Cyclone," Air Weather Service, Tech. Report 164, August 1962.

[11] Brooks, C. E. P. and Carruthers, N.: Handbook of Statistical Methods in Meteorology, Her Majesty's Stationary Office, London, 1953.

[12] Brooks, C. E. P. and Evans, G. T.: "Annotated Bibliography on Industrial Meteorology," Meteorol. Abst. and Bibliogr., Vol. 5, No. 10, October 1954, pp. 1187-1230.

[13] Brown, D. R. E.: "Natural Illumination Charts," Bureau of Ships, U.S. Navy, Washington, D. C., 1952.

[14] Charles, B. N.: "Empirical Models of Interlevel Correlation of Winds," J. Meteorol., Vol. 16, No. 5, October 1959, pp. 581-585.

[15] Charles, B. N.: "Utility of Stretch Vector Correlation Coefficients," Quart. J. Roy. Meteorol. Soc., Vol. 85, No. 365, 1959, pp. 287-290.

[16] Conrad, V. A. and Pollack, L. W.: Methods in Climatology, 2d Edition, Harvard Univ. Press, 1950.

[17] Court, A.: "Some New Statistical Techniques in Geophysics," Advan. Geophys., Vol. 1, 1952, pp. 45-85.

[18] Court, A.: "Wind Correlation and Regression," AFCRC TN-58-230, 19 February 1958 (AD-152460).

[19] Crutcher, H. L.: "On the Standard Vector-Deviation Wind Rose," J. Meteorol., Vol. 14, No. 1, February 1957, pp. 28-33.

[20] Cry, G. W., Haggard, W. H., and White, H. S.: "North Atlantic Tropical Cyclones," Tech. Paper No. 36, U.S. Weather Bureau, Washington, D. C. 1959.

[21] Dixon, W. J. and Massey, F. J.: Introduction to Statistical Analysis, 2d Edition, McGraw-Hill Book Co., Inc., New York, 1957.

[22] Drozdov, O. A.: Principles of Climatological Analysis of Meteorological Observations, Leningrad, 1956 (In Russian).

[23] Durst, C. S.: "Variation of Wind with Time and Distance," Geophys. Memoirs No. 93, Meteorological Office, Air Ministry, London, 1954.

[24] Goldberg, Samuel: Probability: An Introduction, Prentice Hall, 1960.

[25] Goodman, R.: Modern Statistics, Arco Publishing Co., Inc., New York, 1964.

[26] Gringorten, I. I.: "Extreme Value Statistics in Meteorology — a Method of Application," AF Surveys in Geophysics No. 125, June 1960.

[27] Gumbel, E. J.: Statistics of Extremes, Columbia Univ. Press, New York, 1958.

[28] Harter, H. L.: "Circular Error Probabilities," J. Am. Statistical Assoc., Vol. 55, No. 292, December 1960, pp. 723-731.

[29] Jacobs, W. C.: "Wartime Developments in Applied Climatology," Meteorol. Monographs, AMS, Vol. 1, No. 1, August 1947.

[30] Kochanski, A.: "Models of Vertical Correlations of the Wind," J. Meteorol., Vol. 18, No. 2, April 1961, pp. 151-159.

[31] Landsberg, H. E. and Jacobs, W. C.: "Applied Climatology," in: Compendium of Meteorology, AMS, Boston, Massachusetts, August 1951, pp. 976-992.

[32] Lieblein, Julius: "A New Method of Analyzing Extreme-Value Data," NACA Tech. Note 3053, Washington, D. C., January 1954.

[33] McCabe, J. T.: "Climatology of Far East Typhoons, a Preliminary Report," Hq, 10th Weather Group, Air Weather Service, January 1960.

[34] McMurray, W. M.: "Data Collection for the Northern Hemisphere Map Series," Monthly Weather Rev., Vol. 84, No. 6, June 1956, pp. 219-234.

[35] Malone, T. F.: "A Method for Preparing Probability Forecasts Applicable to Decision-Making When a Hurricane Threatens," Travelers Insurance Co., Hartford, Connecticut, 1959.

[36] Mauchly, J. W.: "A Significance-Test for Ellipticity in the Harmonic Dial," Terrest. Magnetism Atmospheric Elec., Vol. 45, 1940, pp. 145-148.

[37] Mitchell, J. M. Jr.: "Some Practical Considerations in the Analysis of Geophysical Time Series," paper presented at 44th Annual Meeting of AGU, Washington, D. C., April 1963.

[38] Panofsky, H. A. and Brier, G. W.: Some Applications of Statistics to Meteorology, the Pennsylvania State University, University Park, Pennsylvania, 1958.

[39] Quiroz, R. S.: "A Note on Meteorological Glossaries," Bull. Amer. Meteorol. Soc., Vol. 39, No. 5, May 1958, pp. 282-283.

[40] Quiroz, R. S.: "A Survey of Upper Air Climatological Information Sources," Air Weather Service, Climatic Center TM 59-4, September 1959.

[41] Rigby, M. and Gropp, D. M.: "Bibliography of Bibliographies in Meteorology," Meteorol. Abstr. and Bibliogr., Vol. 6, No. 1, January 1955, pp. 82-118.

[42] Sawyer, J. S.: "Equivalent Headwinds; Application of Upper-Wind Statistics to Air-Route Planning," Great Britain Meteorological Office, Meteorol. Report No. 6, M.O. 535a, London, 1950.

[43] Scott, J. R.: "A Note on the Parametric Representation of Non-Circular Wind Distributions," Quart. J. Roy. Meteorol. Soc., Vol. 82, No. 351, January 1956, pp. 87-88.

[44] Scott, J. R.: "A Regression Method for Estimating Time Changes of Elliptically Distributed Winds," Quart. J. Roy. Meteorol. Soc., Vol. 82, No. 353, July 1956, pp. 337-339.

[45] Spiegel, M. R.: Theory and Problems of Statistics, Schaum's Outline Series, Schaum Publishing Co., New York 1962.

[46] Thom, H. C. S.: "Distributions of Extreme Winds in the United States," Journal of the Structural Division, Proc. Am. Soc. of Civil Eng., Vol. 86, No. ST4, April 1960.

[47] Thomson, J. C. and Brier, G. W.: "The Economic Utility of Weather Forecasts," Monthly Weather Rev., Vol. 83, No. 11, November 1955, pp. 249-254.

[48] Thornthwaite, C. W. and Mather, J. R.: "Instructions and Tables for Computing Potential Evapotranspiration and the Water Balance," Publications in Climatology, Vol. X, No. 3, Laboratory of Climatology, Drexel Institute, 1957.

[49] Thornthwaite, C. W. and Mather, J. R.: "The Water Balance," Publications in Climatology, Vol. VIII, No. 1, Laboratory of Climatology, Drexel Institute, 1955.

[50] Thornthwaite, C. W., et al.: "Estimating Soil Moisture and Tractionability Conditions for Strategic Planning," AF Surveys in Geophysics No. 94, 1958 (AD-146789).

[51] USSR: Meteorologicheskii Biulleten' (for July 1957-December 1958), Tsentr. Inst. Prognozov, 1958-1959.

[52] U.S. Office of Naval Operations: U.S. Navy Marine Climatic Atlas of the World, Vols. 1-5, 1955-1959.

[53] U.S. Weather Bureau: "Atlas of Climatic Charts of the Ocean," 1938.

[54] U.S. Weather Bureau: "10-Millibar Synoptic Weather Maps -- July 1957-June 1958, of the IGY Period," December 1959.

[55] Yule, G. U. and Kendall, M. G.: An Introduction to the Theory of Statistics, Hafner Publishing Co., New York, 1950, pp. 1-68.

HQ AIR WEATHER SERVICE OFFICIAL U S AIR FORCE

RUSSELL K. PIERCE, JR., Brigadier General, USAF
Commander

CHARLES W. TIEMANN
Director of Administration

DISTRIBUTION: X (Continued)

Hq AWS	50
Wings	5
except 7 Wea Wg	17
USAF ETAC	50
Squadrons	2
Detachments	1
Special	

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D		
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)		
1. ORIGINATING ACTIVITY (Corporate author) USAF Environmental Technical Applications Center, Bldg. 159, Navy Yard Annex Washington, D. C. 20333	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/A	
3. REPORT TITLE Guide for Applied Climatology		
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) N/A		
5. AUTHOR(S) (First name, middle initial, last name) N/A		
6. REPORT DATE November 1968	7a. TOTAL NO. OF PAGES 168	7b. NO. OF REFS 55
8a. CONTRACT OR GRANT NO. N/A b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Air Weather Service Pamphlet 105-2 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) N/A	
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale; its distribution is unlimited.		
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Hq Air Weather Service (MAC) Scott AFB, Illinois 62225	

13. ABSTRACT

This pamphlet explores the functions of an applied climatological unit especially orientated towards, but not limited to, support of the military. The "data base" available within the Environmental Technical Applications Center (ETAC) is explained in detail. Statistical probabilistic and empirical applications to climatic procedures are illustrated and several examples of problem-solving techniques are included.

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Data base Climatological techniques Wind distribution Statistics Probabilities						

IED

-78